Fundamental Limits on Orbit Uncertainty

Daniel J. Scheeres
Department of Aerospace Engineering Sciences
University of Colorado
Boulder, Colorado 80309
scheeres@colorado.edu

Maurice A. de Gosson
Fakultät für Mathematik
Universität Wien, NuHAG
A-1090 Vienna

Jared M. Maruskin
Department of Mathematics
San José State University
San José, California 95192

***Abstract*—The orbital environment of Resident Space Objects (RSO) is highly structured and deterministic in general, with stochastic effects only arising due to mis-modeling and to some variability in the physical environment in which they orbit. Due to this RSO can be well modeled as following Hamiltonian Dynamics, and thus their state space can be expressed as motion on a symplectic manifold, including their uncertainty distributions. There has been recent progress in understanding the fundamental dynamics of phase flows in symplectic systems, with the preeminent advance being "Gromov's Non-Squeezing Theorem," which can be shown to be similar to the famous Heisenberg Uncertainty Principle in quantum mechanics. This paper applies this concept to probability distribution functions and discusses the resulting implications for their dynamics and measurement updates.**

I. Introduction

Motion of Resident Space Objects (RSO) in many cases can be modeled using Hamiltonian Dynamics, due to the detail with which the physics acting on them are known and the relative weakness of the non-conservative perturbations (for objects not in Low Earth Orbit). Thus, the dynamical mapping of RSO orbit solutions and their attendant probability density functions can be shown to be symplectic maps, and thus must conform to the constraints of symplectic topology. Recent advances in this field [1], [2] have discovered fundamental, and unsuspected, constraints on the mapping of finite volumes of phase space – which are at the heart of dynamical mappings for orbit uncertainty.

To motivate a statement of Gromov's non-squeezing theorem, consider a closed Euclidean ball $B^{2n}(r)$ in $\mathbb{R}^{2n}$ centered at 0 with radius r, and a symplectic cylinder Z^{2n} defined as

$$Z^{2n}(r) \quad = \quad B^2(r) \times \mathbb{R}^{2n-2}$$

We should note that the disc $B^2(r)$ is symplectic, which means its coordinates are a coordiante/momentum pair (q_i, p_i) with the constraint $q_i^2 + p_i^2 \leq r^2$. Let $\varphi : U \to V$ be a symplectic embedding, where $U, V \subset \mathbb{R}^{2n}$. Then the non-squeezing theorem gives a constraint on the action of φ on the phase volume mapping:

Non-squeezing Theorem: If there is a symplectic embedding $B^{2n}(r) \hookrightarrow Z^{2n}(R)$, then $r \leq R$.

In [3], [4], [5], [6] applications of Gromov's non-squeezing theorem are made to the study of uncertainty analysis of Hamiltonian systems. These papers study the implications of the non-squeezing theorem for probability density functions that define uncertainty distributions for particle trajectories in space, with specific applications to spacecraft [3], [6] and orbit debris predictions [5]. In addition, Gromov's theorem is used to develop a constraint on covariance matrices for linear Hamiltonian systems of the form used in spacecraft navigation and orbit determination [4].

This paper reviews and combines these previous studies with a focus on systematically using the notion of symplectic capacity, and shows that these results can be restated in a form very similar to the strong uncertainty principle of quantum mechanics due to Robertson [7] and Schrödinger [8]. In addition, we derive formulae to update the symplectic width of a Hamiltonian Dynamical System's probability density function using Bayes' Theorem, thus showing how these concepts can be generalized to navigation problems, potentially making the Gromov width of a distribution into a metric for evaluating the quality of knowledge of a state. In our analysis we find it important to distinguish between two types of probability distribution functions, the classical distributions which are non-zero over all of phase space and an alternate form which have a compact basis, meaning that they are only non-zero over a finite volume in phase space.

II. Notation and terminology

A generic point in the phase space of an n degree of freedom system is denoted as $X = (q, p) \in \mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}$, where we have set $q = (q_1, ..., q_n)$, $p = (p_1, ..., p_n)$. When matrix calculations are performed, X, p, q are always viewed as column vectors.

The symplectic group of $\mathbb{R}^{2n}$ is denoted by $\mathrm{Sp}(n)$ and is the group of all real $2n \times 2n$ matrices S such that $S^T J S = J$ (or, equivalently, $S J S^T = J$) where

$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

and 0 and I are the $n \times n$ zero and identity matrices, respectively.

With the above definitions we state that a Hamiltonian Dynamical System is defined as a set of $2n$ differential equations defined using a Hamiltonian function $H(X, t)$ according to the rule:

$$\dot{X} = J \frac{\partial H}{\partial X}$$

The solution flow is a one-parameter group of symplectomorphisms and thus represent volume-preserving mappings of phase space from some initial distribution to a final distribution. When these distributions have non-zero volume they will be subject to Gromov's Theorem.

III. Statistical Description of Hamiltonian Dynamical Systems

A. Probability Density Functions

The uncertainty of a Hamiltonian state X, such as computed when tracking a spacecraft, can be characterized using a probability density function (pdf), denoted as $\rho(t, X)$. Given an initial pdf, $\rho(t_o, X)$, we obtain a one-parameter family of pdf's, parameterized by time. A probability density function $\rho : \mathbb{R} \times \mathbb{R}^{2n} \to \mathbb{R}$ is a function that satisfies the properties:

- $\rho(t, X) \geq 0$ for $X \in \mathbb{R}^{2n}$.
- $\int_{\mathbb{R}^{2n}} \rho(t, X) dX = 1$.

The pdf forms the basis of spacecraft navigation, and it generates the most important quantities for these systems. The main statistical application of the pdf is found by interpreting this function as a generator of the probability of finding the object within a certain volume of phase space. Given a set in phase space, $\mathcal{B} \subset \mathbb{R}^{2n}$, with a non-zero volume, $\mathrm{Vol}_{2n}(\mathcal{B}) > 0$, the probability of the state lying in this set is simply computed as the $2n$-dimensional integral of the pdf over the set $\mathcal{B}$,

$$P(\mathcal{B}) = \int_{\mathcal{B}} \rho(X) dX \tag{1}$$

We define

$$\mathrm{Vol}_{2n}(\mathcal{B}) = \int_{\mathcal{B}} dX$$

to represent the $2n$ dimensional volume of the set $\mathcal{B}$ in phase space. We note that there can be an infinite number of sets that may have a given probability, although the probability of each set has a characteristic value.

Given a pdf, the major interest of classical spacecraft navigation is in the first few moments of that distribution. The *first moment* generates the mean of the state, which is generally taken to represent the most likely solution. The *second moment* generates the covariance of the state, which is taken to represent the relative uncertainties in the state. Higher order moments can also be defined, but are not used that frequently in orbital mechanics. Current topics of interest in the field of astrodynamics, however, have focused back on the fundamental measurement and state pdf, motivating our more general discussion of this topic.

There are three key operations that one may perform on the pdf, which can change the mean and covariance of the states. The first fundamental operation is that of dynamical propagation. In this work we disregard the stochastic effect of accelerations on a dynamical system for simplicity, and for definiteness as such additional effects can destroy the symplectic properties of the dynamics. In the absence of such stochastic effects the pdf is conserved at each evolving point in phase space.

We can define ρ as a solution of an initial value problem for PDE's; i.e., given $\rho_0(t, X)$, $\rho : \mathbb{R}^{2n} \times \mathbb{R} \to \mathbb{R}$ is the solution of a PDE IVP

$$\frac{D\rho}{Dt} = \frac{\partial \rho}{\partial t} + \dot{x} \cdot \nabla \rho = 0, \qquad \rho(X) = \rho_0(X).$$

We will suppress the time dependence in the following. Thus we note that the total time rate of change of the pdf is zero under our deterministic dynamical assumptions, implying that the probability density function is conserved along each trajectory for a Hamiltonian system. Due to this fact the probability of a set is an integral invariant. Thus, even though the set itself will propagate over time, the value of probability associated with a set is constant unless some other operation is applied to the pdf.

Whereas the pdf is conserved along the symplectic flow, which represents dynamical propagation of the system state, it may change discontinuously, at a moment in time, following a physical measurement of the system. We define an associated probability density function for a measurement of a state as a function of an abstract observation, $z \in \mathbb{R}$ and a measurement function $H(X) : \mathbb{R}^{2n} \to \mathbb{R}$. In the absence of noise we have a correspondence between the observation of a state and the measurement function: $H(X) = z$. In real systems, there is additional uncertainty associated with the given observation, may also be described with a pdf. Usual assumptions are that the uncertainty has zero mean relative to the true observation, and that the most important aspect of the uncertainty is described by its second moment. We will simply describe a given observation by a measurement pdf, denoted as $\rho_H(X)$, and note that it has a mean, variance, and other specific structures.

$$z = \int_{\infty} H(X) \rho_H(X) dX \tag{2}$$

$$\sigma^2 = \int_{\infty} (H(X) - z)^2 \rho_H(X) dX \tag{3}$$

Other modifications can be added, such as the inclusion of unknown and statistically defined biases in the measurement pdf.

We use a simple Bayesian update rule for incorporating measurement information into the state pdf:

$$\rho'(X) = \frac{\rho(X)\rho_H(X)}{\int_{\infty} \rho(X)\rho_H(X) dX} \tag{4}$$

Finally, we are also interested in computing conditional pdfs, which generates a pdf exclusively for a sub-state. Assume we split our general state $X = [x_1, x_2, \ldots, x_j, \ldots, x_n]$, where $x_i = (q_i, p_i)$ and forms a symplectic pair. The marginal pdf of x_i is then:

$$\rho(x_i) = \int_{\infty} \rho(X) dx_{j \neq i} \tag{5}$$

where the integral occurs over all states other than x_i. We can arbitrarily choose different combinations of which states to integrate over and which to remain free.

Most classical pdf descriptions are non-zero over the entire $\mathbb{R}^{2n}$ space. Thus the volume to generate a unity probability is infinite for this class of pdf. An alternate, and in some sense more realistic, approach is to define the pdf to be non-zero only over a volume of finite measure, meaning that ρ has compact support. This respects the obvious fact that an object does not have a probability to exist at all points in space, but in general is usually strongly localized.

We will need to distinguish between these two types of pdfs in the following, classical distributions that have non-zero probability over all space and finite volume distributions that only have non-zero probabilities over limited regions of space. Just as with state pdfs, measurement pdfs can also be defined to be non-zero over a finite volume set (e.g., [9]) or, what is more usual, be defined over all space.

B. Classical Normal Distributions

We will exclusively use the Gaussian, or Normal, distribution as a representative of the class of distributions which have non-zero probability over all space. For space applications, the Gaussian pdf is by far the most applicable and popular. We provide a representation of this pdf as:

$$\rho(X) = \frac{1}{(2\pi)^n \sqrt{|P|}} \exp\left[-\frac{1}{2}\left(X - \bar{X}\right)\Lambda\left(X - \bar{X}\right)\right] \quad (6)$$

where $\Lambda = P^{-1}$ is defined as the information matrix and P is the covariance matrix. The dynamic evolution of a Gaussian pdf, in general, will not remain Gaussian (except in the restricted case of a linear dynamical system). Thus a non-linear dynamics mapping of the state will destroy the Gaussian form in general, although it can still be used to describe the initial conditions of the pdf (e.g., [10]).

To deal with sets of finite volume, we must restrict ourselves to sets that have less than full probability, meaning that all of our applications will not address all possible states, but only a subset of the possible states. Again, there are arbitrary ways in which we can choose our sets, but the most meaningful and descriptive approach chooses states that have a density greater than or equal to some value. These are obviously defined as ellipsoids via:

$$\mathcal{E}_r = \left\{X | \left(X - \bar{X}\right)\Lambda\left(X - \bar{X}\right) \leq r^2\right\} \quad (7)$$

It is a classical computation to find the probability associated with such a set.

C. Finite Volume Distributions

The pdfs for finite volume distributions are defined such that $\rho(X)$ takes on all of its positive values over a finite volume of phase space. These are not commonly used, but from a physically motivated perspective are more realistic than the above Gaussian distribution which retains non-zero probability over all of phase space. The probability outside the intersection of these finite sets is automatically ruled out; this fact makes finite volume distributions computationally superior, as one no longer need integrate over the all of space. Despite these advantages, the level of definiteness which we can bring to the discussion in this case decreases greatly, due to the lack of a generally accepted analytical form for these pdfs. Thus, in the following we will only deal with their abstract properties.

In the following we stipulate that the set for which the pdf takes on non-zero values is a connected, compact set in phase space. We will denote this minimal set as $\mathcal{B}$ and note that it will have a volume $V_{\mathcal{B}} = \text{Vol}_{2n}(\mathcal{B}) < \infty$ and that $\int_{\mathcal{B}} \rho(X)dX = 1$. In addition to these, we will also assume that $\rho(X) = 0$ if $X \notin \mathcal{B}$ (note that this eliminates sets of measure zero where $\rho(X)$ is non-zero).

Given this definition of a finite volume set, we can introduce a more conservative parameterization of this pdf by introducing an ellipsoid, denoted by $\mathcal{E}_{\mathcal{B}}$, which has a finite volume and which contains the set $\mathcal{B}$. Selection of this ellipsoid is not unique, and we do not consider how it is constructed. We note that $\rho(X) \geq 0$ over the set $\mathcal{E}_{\mathcal{B}}$ and that $\int_{\mathcal{E}_{\mathcal{B}}} \rho(X)dX = 1$ as before. Selection of this bounding ellipsoid is conservative in the sense that the volume of this ellipsoid will always be greater than or equal to the volume of the original set: $\text{Vol}_{2n}(\mathcal{B}) \leq \text{Vol}_{2n}(\mathcal{E}_{\mathcal{B}})$. The consequences of this conservative bound will be seen later when we discuss the Gromov width of our distributions. However, in other situations, it is a convenient way to introduce analytical constraints on these finite volume pdf distributions. An alternate approach is to construct finite-volume distributions with cut-off Gaussians: Renormalized Gaussian functions set to zero outside a certain confidence ellipsoid. In such a case, the "conservative" ellipsoid is actually tight.

IV. Symplectic Capacities and the Gromov Width of Distributions

Gromov's theorem can be used to develop a constraint on covariance matrices for linear Hamiltonian systems of the form used in spacecraft navigation and orbit determination. In particular, it is proven in [4], Theorem 1, p.688, that the symplectic width $w_L(\mathcal{E})$ of the Gaussian uncertainty ellipsoid $\mathcal{E} : X^T P^{-1} X \leq 1$ is related to the covariance matrix P_{ii} in the q_i, p_i variables by the inequality

$$|P_{ii}| \geq \left(\frac{w_L(\mathcal{E})}{\pi}\right)^2. \quad (8)$$

where $|P_{ii}|$ is the determinant of the covariance matrix projected onto the sub-state $x_i = (q_i, p_i)$.

In the following subsections we revisit this earlier calculation and show that it can be derived using alternate approaches that have a deeper connection with the notion of the Heisenberg Uncertainty Principle and Quantum Mechanics in general. We first define and describe the concept of symplectic capacities and the Gromov width of a distribution in phase space.

A. The symplectic capacity of ellipsoids

Gromov's theorem can be restated as follows: Let $\mathcal{P}_j$ be a symplectic plane (for instance any plane of coordinates q_j,p_j. The projection of the ball $B^{2n}(r)$ on $\mathcal{P}_j$ is a circle with area πr^2. Let us now apply a canonical transformation Φ to $B^{2n}(r)$, it becomes a new phase space set $\Phi(B^{2n}(r))$ which can have a very different shape from the origin ball (it will however have the same volume, in view of Liouville's theorem). Gromov's theorem tells us that the projection of the distorted ball $\Phi(B^{2n}(r))$ on the symplectic plane $\mathcal{P}_j$ will have area at least πr^2.

All known proofs of Gromov's theorem are notoriously difficult, which may explain why the non-squeezing property wasn't known before the mid 1980's. In our context the important thing is that this theorem makes it possible to define a new geometric notion, that of the symplectic width w_L of an arbitrary subset of phase space. Such a symplectic width is constructed as follows: Let Ω be a set in $\mathbb{R}^{2n}$. Assume first that there is no canonical transformation mapping a phase-space ball $B^{2n}(r)$ inside Ω, no matter how small its radius r is. We will then write $w_L(\Omega) = 0$. (A typical example of this situation is when Ω has dimension smaller than $2n$ as follows from Liouville's theorem.) Assume next that we can find at least one canonical transformation sending some ball $B^{2n}(r)$ within this set Ω. The supremum R of all such radii r is called the *symplectic radius* of Ω and we define the symplectic width by $w_L(\Omega) = \pi R^2$. The symplectic width has the three following properties:

$$w_L(\Omega') \leq w_L(\Omega) \textit{ if } \Omega' \textit{ is a subset of } \Omega; \tag{9a}$$

$$w_L(f(\Omega)) = w_L(\Omega) \textit{ if } f \textit{ is a canonical transformation}; \tag{9b}$$

$$w_L(\lambda\Omega) = \lambda^2 w_L(\Omega) \textit{ for every real number } \lambda. \tag{9c}$$

It has in addition the following fundamental property:

$$w_L(B^{2n}(r)) = \pi r^2 = w_L(Z_j^{2n}(r)). \tag{10}$$

While the three first properties listed above are satisfied by many functions (for instance $(\mathrm{Vol}_{2n}(\Omega))^{1/n}$) where $\mathrm{Vol}_{2n}(\Omega)$ is the volume of Ω if Ω is measurable), the fourth property is highly non-trivial, because its proof requires the non-squeezing theorem and is actually equivalent to it. More generally, one calls *symplectic capacity* any function c assigning to a subset Ω a finite or infinite number $c(\Omega)$ such that properties (9)–(10) are satisfied (Ekeland and Hofer [2]). It turns out that there exist infinitely many symplectic capacities, and that the symplectic width w_L is the *smallest* of all symplectic capacities. A very interesting symplectic capacity c_{HZ} was constructed by Hofer and Zehnder in [11]. It has the property that when Ω is a bounded and convex set in phase space then we have

$$c_{\mathrm{HZ}}(\Omega) = \oint_{\gamma_{\min}} pdq \tag{11}$$

where $pdq = p_1 dq_1 + \cdots + p_n dq_n$ and $\gamma_{\min}$ is the shortest (positively oriented) Hamiltonian periodic orbit carried by the boundary $\partial\Omega$ of Ω.

A remarkable fact is that all symplectic capacities agree on phase space ellipsoids. Let us describe this fact more in detail. Let P be a positive definite real $2n \times 2n$ matrix (and can be considered to be equivalent to the covariance of our distribution); the matrix JP being equivalent to the antisymmetric matrix $P^{1/2}JP^{1/2}$ the eigenvalues of JP are of the type $\pm i\lambda_j$ with $\lambda_j > 0$. The positive numbers λ_j are called the "symplectic eigenvalues" of P. We will always arrange these numbers λ_j in *decreasing* order:

$$\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = \lambda_{\min} \tag{12}$$

and call the sequence

$$\mathrm{Spec}_{\mathrm{Sp}}(P) = (\lambda_1, ..., \lambda_n)$$

the *symplectic spectrum* of P. Since JP^{-1} has the same eigenvalues as $(JP^{-1})^{-1}$ we have

$$\mathrm{Spec}_{\mathrm{Sp}}(P^{-1}) = (\lambda_n^{-1}, ..., \lambda_1^{-1}).$$

Proposition 1: Let Ω^+ and Ω^- be the phase space ellipsoids defined by $X^T PX \leq 1$ and $X^T P^{-1} X \leq 1$, respectively. For every symplectic capacity c we have

$$c(\Omega^+) = \frac{\pi}{\lambda_{\max}} \quad , \quad c(\Omega^-) = \pi\lambda_{\min}. \tag{13}$$

Proof: See e.g. [11]. ■

A classical result, due to Williamson [12], is the following: there exists $S \in \mathrm{Sp}(n)$ such that

$$S^T PS = \begin{bmatrix} \Lambda & 0 \\ 0 & \Lambda \end{bmatrix} \tag{14}$$

where $\Lambda = \mathrm{diag}\,[\lambda_1, ..., \lambda_n]$ (see for instance Hofer and Zehnder [11] for a modern proof). It follows that the transforms of the ellipsoids Ω^+ and Ω^- are represented by, respectively

$$\sum_{j=1}^{n} \lambda_j (q_j^2 + p_j^2) \leq 1 \text{ and } \sum_{j=1}^{n} \lambda_j^{-1} (q_j^2 + p_j^2) \leq 1.$$

B. The ε-Condition

Let P be a positive definite real $2n \times 2n$ matrix; we will make extensive use of the following condition on P:

> *If there exists an $\varepsilon > 0$, such that the eigenvalues of the complex Hermitian matrix $P + i\varepsilon J$ are all non-negative (which we write for short as $P + i\varepsilon J \geq 0$) we will say that P "satisfies the ε-condition".*

That $P + i\varepsilon J$ is Hermitian follows from the relations $P^* = P^T = P$ and $(iJ)^* = (-i)J^T = iJ$. To say that P satisfies the ε-condition is equivalent to saying that

$$Z^*(P + i\varepsilon J)Z \geq 0 \tag{15}$$

for all complex vectors $Z = X + iY$ in $\mathbb{C}^{2n}$. We observe that if P satisfies the ε-condition then it also satisfies the ε'-condition for every $\varepsilon' \leq \varepsilon$. To see this, set $\varepsilon' = r\varepsilon$ with $0 < r \leq 1$. We have

$$P + i\varepsilon' J = (1 - r)P + r(P + i\varepsilon J)$$

hence $P + i\varepsilon' J \geq 0$ since $(1 - r)P \geq 0$ and $r(P + i\varepsilon J) \geq 0$.

Theorem 2: Let Ω^- be the ellipsoid $X^T P^{-1} X \leq 1$. The matrix P satisfies the ε-condition for every $\varepsilon \leq \frac{1}{\pi} c(\Omega^-)$. In particular

$$P + \frac{i}{\pi} c(\Omega^-) J \geq 0. \tag{16}$$

Proof: Let S be a symplectic matrix diagonalizing P following formula (14). We have

$$S^T(P+i\varepsilon J)S = S^TPS + i\varepsilon S^TJS = S^TPS + i\varepsilon J$$

and hence

$$S^T(P+i\varepsilon J)S = \begin{bmatrix} \Lambda & i\varepsilon I \\ -i\varepsilon I & \Lambda \end{bmatrix}.$$

The condition $P+i\varepsilon J \geq 0$ is equivalent to $S^T(P+i\varepsilon J)S \geq 0$; using the equality above this condition is easily seen to be equivalent to $\lambda_j^2 - \varepsilon^2 \geq 0$ for $1 \leq j \leq n$, that is to $\lambda_j \geq \varepsilon$ since the λ_j are nonnegative. It follows that

$$c(\Omega^-) = \pi\lambda_n \geq \pi\varepsilon$$

in view of the second formula (13). The result follows. ∎

C. A Classical Uncertainty Principle

Assume that the orbit distribution is initially given by a multivariate Gaussian distribution with covariance matrix P and average $\bar{X}$:

$$\rho_0(X) = \frac{1}{(2\pi)^n\sqrt{|P|}} \exp\left[-\frac{1}{2}(X-\bar{X})P^{-1}(X-\bar{X})\right].$$

To that distribution we associate the uncertainty ellipsoid

$$\mathcal{E}_0 : \frac{1}{2}(X-\bar{X})P^{-1}(X-\bar{X}) \leq 1.$$

Applying the second formula (13) to P the symplectic capacity of $\mathcal{E}_0$ is

$$c(\mathcal{E}_0) = \pi\lambda_{\min}^P$$

where $\lambda_{\min}^P$ is the smallest symplectic eigenvalue of P.

Rearranging the coordinates (q,p) as $(q_1,p_1,q_2,p_2,...,q_n,p_n)$ we can write the covariance matrix as

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

where the P_{ij} are the symmetric 2×2 blocks

$$P_{ij} = \begin{bmatrix} \sigma_{q_iq_j} & \sigma_{q_ip_j} \\ \sigma_{q_ip_j} & \sigma_{p_ip_j} \end{bmatrix};$$

the quantities $\sigma_{q_iq_j}$, $\sigma_{q_ip_j}$, $\sigma_{p_ip_j}$ are the covariances in the indexing variables.

As already noticed in [3], in usual navigation practice the navigation performance of a spacecraft is specified as a function of the state on some lower dimensional subspace on which probabilities are computed. We restate this in the following form:

Theorem 3: Assume that P satisfies the ε-condition.
(i) Then

$$\sigma_{q_iq_i}\sigma_{p_ip_i} \geq (\sigma_{q_ip_i})^2 + \varepsilon^2 \quad \text{for} \quad i = 1,2,...,n. \tag{17}$$

(ii) In particular we recover the Hsiao–Scheeres formula

$$|P_{ii}| \geq \left(\frac{c(\mathcal{E}_0)}{\pi}\right)^2.$$

Proof: The non-negativity of the Hermitian matrix $P + i\varepsilon J$ implies that

$$\begin{bmatrix} \sigma_{q_iq_i} & \sigma_{q_ip_i} + i\varepsilon \\ \sigma_{q_ip_i} - i\varepsilon & \sigma_{p_ip_i} \end{bmatrix} \geq 0$$

(see de Gosson [13]); since $\sigma_{q_iq_j} \geq 0$ and $\sigma_{p_ip_j} \geq 0$ this is equivalent to

$$\sigma_{q_iq_i}\sigma_{p_ip_i} - (\sigma_{q_ip_i} + i\varepsilon)(\sigma_{q_ip_i} - i\varepsilon) \geq 0$$

which is the same thing as (17). The statement (ii) follows from (17) and Theorem 2. ∎

D. Propagation of Uncertainty: the Linear Case

Consider as in Hsiao and Scheeres [4] the dynamical system

$$\dot{X}(t) = A(t)X(t) \quad , \quad X(t_0) = X_0. \tag{18}$$

We write its solution in the form $X(t) = \Phi(t,t_0)X_0$, since the system is linear the operator $\Phi(t,t_0)$ is also linear and can be identified with a $2n \times 2n$ matrix. In what follows we assume that the product $JA(t)$ is symmetric (which always follows for a Hamiltonian linear system). It follows that $A(t)$ belongs to the Lie algebra of $\mathrm{Sp}(n)$ and that (18) is thus equivalent to the Hamiltonian system

$$\dot{X}(t) = J\frac{\partial H}{\partial X}(X(t),t)$$

for the special case in which H is the (time-dependent) Hamiltonian function

$$H(X,t) = -\frac{1}{2}X^T(JA(t))X.$$

The matrices $\Phi(t,t_0)$ are therefore symplectic for all t and t_0.

We denote by P_0 the covariance matrix at initial time t_0. Then the covariance matrix at arbitrary time t is

$$P(t) = \Phi(t,t_0)P_0\Phi(t,t_0)^T.$$

Theorem 4: Assume that P_0 satisfies the ε-condition. Then $P(t)$ also satisfies the ε-condition. In particular

$$\sigma_{q_iq_i}(t)\sigma_{p_ip_i}(t) \geq (\sigma_{q_ip_i}(t))^2 + \varepsilon^2 \quad \text{for} \quad i = 1,2,...,n.$$

Proof: It immediately follows from the observation that

$$P(t) + i\varepsilon J = \Phi(t,t_0)(P_0 + i\varepsilon J)\Phi(t,t_0)^T$$

since $\Phi(t,t_0)J\Phi(t,t_0)^T = J$ because $\Phi(t,t_0)$ is symplectic. ∎

E. The case of completely integrable systems

We add the following observations as a counterpoint and generalizations of our results for ellipsoids. As many important systems can be described as integrable systems, it is relevant to discuss the application of Gromov's Theorem to these cases as well.

Recall that a physical system with Hamiltonian function H is called integrable (or Liouville integrable, or completely integrable) if there exists a canonical transformation $(q,p) = \Phi(I,\theta)$ such that $H(q,p) = H(\Phi(I,\theta)) = K(I)$. This is equivalent to being able to obtain a set of n functions that are pairwise in involution with respect to the Poisson bracket. In particular, the Keplerian 2-body problem falls into this class. In this case the Hamilton equations reduce to the simple system

$$\frac{d\theta_j}{dt} = \frac{\partial K}{\partial I_j} \quad , \quad \frac{dI_j}{dt} = \frac{\partial K}{\partial \theta_j} = 0$$

whose solutions are

$$\theta_j(t) = \omega_j(I_j(t_0))(t - t_0) + \theta_j(t_0) \quad , \quad I_j(t) = I_j(t_0). \quad (19)$$

Recall that these are the action-angle variables and topologically represent motion on an n-torus.

Let us now introduce the symplectic polar coordinates

$$Q_j = \sqrt{2I_j} \cos\theta_j \quad , \quad P_j = \sqrt{2I_j} \sin\theta_j;$$

one readily verifies that the Jacobian matrix of the transformation $\Psi(I, \theta) = (Q, P)$ is symplectic, hence Ψ is a canonical transformation (we remark that in contrast the passage to usual polar coordinates is not canonical). In the coordinates (Q, P) the solutions (19) become

$$Q_j(t) = \sqrt{2I_j(t_0)} \cos\theta_j(t) \quad , \quad P_j(t) = \sqrt{2I_j(t_0)} \sin\theta_j(t)$$

and we have $Q_j(t)^2 + P_j(t)^2 = 2I_j(t_0)$ hence, in these coordinates, the motion takes place on the torus

$$T^n = S_1^1(R_1) \times \cdots \times S_n^1(R_n)$$

where $S_j^1(R_j)$ is the circle with radius R_j in the q_j, p_j plane and $R_j = \sqrt{2I_j(t_0)}$. Let us now consider the associated polydisk

$$D^n = D_1^1(R_1) \times \cdots \times D_n^1(R_n)$$

where $D_j^1(R_j)$ is the polydisk ${Q_j}^2 + {P_j}^2 \leq R_j$. One proves (Polterovich [14], Theorem1.1C and Section 4.3) that the symplectic width of that polydisk is

$$w_L(D^n) = \pi R^2 \quad , \quad R = \min_{1 \leq j \leq n} R_j.$$

Assume for clarity that $R = R_1$. In view of Gromov's non-squeezing theorem and the definition of the symplectic width, the supremum of the radii of all phase space balls we can squeeze inside D^n is precisely R_1. Returning to the original coordinates (q, p) it follows, in view of the canonical invariance of symplectic capacities and the fact that we have only used canonical changes of variables, that at any moment the area enclosed by the projection of the trajectory passing through (q_0, p_0) at initial time t_0 on the $x_1 = (q_1, p_1)$ plane is at least $\pi R^2 = 2\pi I_j(t_0)$.

V. Definitions and Modifications of Symplectic Width for Probability Density Functions

Having defined and discussed the symplectic (Gromov) width of a finite set in some detail, we now move on to the second main point of this paper. We wish to understand how to define the symplectic width of a given pdf distribution, or sub-distribution, and then how such a width is modified under the action of dynamics or a state measurement.

We separately consider our two forms of pdfs. First we consider classical normal distribution pdfs. Following that we consider finite volume distribution pdfs.

A. Classical Normal Distributions

Assume again the Gaussian, or Normal, distribution as a representative of the class of distributions which have non-zero probability over all space:

$$\rho(X) = \frac{1}{(2\pi)^n \sqrt{|P|}} \exp\left[-\frac{1}{2} \left(X - \bar{X} \right) \Lambda \left(X - \bar{X} \right) \right] \quad (20)$$

where $\Lambda = P^{-1}$ and is defined as the information matrix. To deal with sets of finite volume, we must restrict ourselves to sets that have less than full probability, meaning that our applications will not address all possible states, but only a subset of the possible sets. Again, there are arbitrary ways in which we can choose our sets, but the most meaningful and descriptive approach chooses states that have a density greater than or equal to some value:

$$\mathcal{E}_r = \left\{ X | \left(X - \bar{X} \right) \Lambda \left(X - \bar{X} \right) \leq r^2 \right\} \quad (21)$$

For definiteness we take $r = 1$ in the following.

Gromov Width Computation: Given an ellipsoid that represents the probability of finding the state within a certain volume, we can apply our previous results to compute the Gromov width of this ellipsoid. Borrowing the concept of the symplectic eigenvalues of the covariance P, defined earlier, we note that the smallest of these, λ_{min}^P, is directly proportional to the Gromov width, or

$$\omega_L(\mathcal{E}_1) = \pi \lambda_{\text{min}}^P$$

For a given momentum coordinate pair with its individual covariance

$$P_{ii} = \begin{bmatrix} \sigma_{q_i q_i} & \sigma_{q_i p_i} \\ \sigma_{q_i p_i} & \sigma_{p_i p_i} \end{bmatrix}$$

The symplectic eigenvalue equals the square root of the determinant, $\lambda_i = \sqrt{\sigma_{q_i q_i} \sigma_{p_i p_i} - \sigma_{q_i p_i}^2}$. Thus, the Gromov width of an ellipsoid distribution in phase space is

$$\omega_L(\mathcal{E}_1) = \pi \min_i |P_{ii}|^{1/2} \quad (22)$$

Thus, in this context the Hsiao-Scheeres formula[4] is trivially derived as $|P_{jj}| \geq \min_i |P_{ii}|$, noting that this inequality is preserved over symplectic, linear mapping of the covariance. The deeper result is that the Gromov width remains invariant under non-linear mapping of the distribution as well.

Mapping in time: We note that the dynamical mapping of a Hamiltonian system in time is always a canonical transformation (equivalently a symplectomorphism). Thus, the symplectic width is conserved through these mappings, whether they are linear or non-linear. If the time mapping is linear, denoted as $X(T) = \Phi(T, t)X(t)$, then the ellipsoidal region will directly map into another ellipsoidal region. Even though the covariance matrix will be altered by the mapping, transforming to $\Phi P \Phi^T$, we note that the symplectic width will remain unaltered. If the time mapping is non-linear, then the ellipsoidal distribution of the original phase volume is lost and becomes distorted. In such cases, the moments of the distribution can be recomputed and a new mean and covariance can be defined (see [10] for a detailed description of this), however the Gromov width associated with this new covariance will in general be different than the previous one. The Gromov width associated with the initial volume distribution associated with the initial pdf will remain invariant, however.

Effects of Measurements: Now consider the effect which a measurement has on the pdf distribution and the corresponding symplectic width. In accordance to the normal distribution which we assume for the state we also assume a standard normal distribution for the measurement pdf, which assumes that the current state is linearized about the current mean. Applying Bayes' rule for the update of the pdf the initial main task is comprised of evaluating the sum of quadratics and re-expressing it in a similar form, leading to the updated information matrix:

$$\Lambda' = \Lambda + \frac{1}{\sigma^2} H_X H_X^T \tag{23}$$

This equation explicitly shows how new information is incorporated into the information matrix, and then as $P' = \Lambda'^{-1}$, how it is incorporated into the covariance matrix. By recursively applying this update, it is possible to increase the information content relevant to a state from zero to a value which allows inversion, assuming that the different H_X form a linear basis.

Thus, resulting from any measurement is an increase in the information (decrease in the covariance) and a shift in the mean. Details on how the nominal solution is used to update the estimate for the next measurement are of interest, but are beyond the current scope of this work. Rather, this result is presented to show how the information matrix, and consequently the Gromov width at a given level of probability, decreases when measurements are obtained. For this classical formulation, there is no theoretical limit to how small the Gromov width can become without introducing relativistic constraints.

Considering the general formula for the update of the information matrix, it is simple to see why this occurs. The new information matrix always consists of the previous matrix summed with a positive semi-definite term of the form $H_X H_X^T$. Thus, the eigenvalues of the information matrix will only grow in general, causing the eigenvalues of the covariance matrix to shrink.

If, instead, we choose to fix the initial volume over which we compute the updated Gromov width then we find that the Gromov width is conserved (trivially) although the action of a measurement will, in general, be to increase the probability of finding the state within that initial volume, although there may be cases where the probability is diluted if the measurement places the object outside of the nominal volume. For these classical distributions, then, we find a simple relation between a decreasing Gromov width for a given probability when measurements are added, or an increasing probability (in most cases) for a fixed Gromov width when measurements are added.

B. Finite Volume Distributions

Now let us assume that the set $\mathcal{B}_1$ has finite volume and $P(\mathcal{B}_1) = 1$, so that the set $\mathcal{B}_1$ contains all possible values of the state, excluding some set of measure zero which we ignore. Assume we can compute the Symplectic Capacity of $\mathcal{B}_1, \omega_L(\mathcal{B}_1)$. An important concept for this class of distribution is the existence of a larger set $\mathcal{B}_1'$ which contains the set $\mathcal{B}_1$, or $\mathcal{B}_1 \subset \mathcal{B}_1'$, allowing for the definition of conservative limits on the set $\mathcal{B}_1$. The choice of $\mathcal{B}_1'$ is arbitrary, meaning that we can choose distributions that are analytically tractable, such as the ellipsoid as discussed above. In this case we know that the full distribution lies within such an ellipsoid and that this object can conservatively stand-in for all possible values of the state. The existence of this larger set will be invoked as convenient in the following discussions.

Gromov width limit: Consider the conditional distribution over the set $x_1 \in \mathcal{B}_1$.

$$\rho_1(x_1) = \int_\infty \rho(X) dx_2 dx_3 \dots dx_n \tag{24}$$

$$= \int_{\mathcal{B}_1} \rho(X) dx_2 dx_3 \dots dx_n \tag{25}$$

The resultant pdf ρ_1 has a finite support over the symplectic plane $x_1 = (q_1, p_1)$, and thus a distribution over this region can be defined, $\mathcal{B}_1^1$. Technically, this is the projection of the entire set onto the x_1 plane, as any region of x_1 that has a non-zero probability over the other planes $x_{j>1}$ is incorporated into this region. Conversely, if there is a coordinate value in the x_1 plane that lies completely outside of all possible values in $\mathcal{B}_1$, then this retains a zero probability and lies outside of the projected set. We assert that the proper interpretation of the Gromov NST is that the area of the projected set, $\mathcal{B}_1^1$, must be greater than or equal to the Gromov width of the original set $\mathcal{B}_1$, or $\mathrm{V}_2(\mathcal{B}_1^1) \geq \omega_L(\mathcal{B}_1)$, which is a further generalization of Eqn. 8, where V_2 represents the 2-dimensional "volume" (area) of the 2-dimensional set B_1^1. The same result holds for any other symplectic plane, of course. Our assertion is made with the caveat that we are referring to the symplectic manifold $(\mathbb{R}^{2n}, \omega_0)$, where ω_0 is the standard symplectic form $\omega_0 = dq \wedge dp$. We note that Gromov's theorem is valid for arbitrary symplectic manifolds (M, ω), which is only *locally* symplectomorphic to $(\mathbb{R}^{2n}, \omega_0)$.

Now consider a set $\mathcal{B}_1'$ defined by an ellipsoid specified as

$$\mathcal{E}_1(\bar{X}, \Lambda) = \left\{ X | (X - \bar{X})^T \Lambda (X - \bar{X}_0) \leq 1 \right\}$$

such that $\mathcal{B}_1 \subset \mathcal{E}_1$. We can also define the projection of this set onto the x_1 plane, and in fact can directly use the result given in Eqn. 22. By the basic properties of symplectic capacities we know that $\mathrm{V}_2(\mathcal{E}_1) = \pi\sqrt{|\Lambda_{11}|} = \omega_L(\mathcal{E}_1) \geq \omega_L(\mathcal{B}_1)$. Additionally we know that $\mathrm{V}_2(\mathcal{E}_1) \geq \mathrm{V}_2(\mathcal{B}_1^1) \geq \omega_L(\mathcal{B}_1)$. However, we do not have specific constraints on the ordinal relation between $\omega_L(\mathcal{E}_1)$ and $\mathrm{V}_2(\mathcal{B}_1^1)$.

Mapping in time: In the current application of mapping entire regions of unity probability, the conservation of the pdf combines with Liouville's Theorem to imply that, while the physical distribution of the set is altered in phase space, its total volume and Gromov width are conserved. Also, a general bounding set $\mathcal{B}_1'$ continues to bound the original set. However, due to non-linear dynamics the bounding ellipsoid $\mathcal{E}_1$ is not conserved as an ellipsoid in general, but will distort into a non-ellipsoidal shape and destroy the simple projection results stated above. Thus, while the concept of a bounding ellipsoid is useful at any given instant, it is not conserved as time progresses except for a linear mapping.

Measurement incorporation: There are a few different situations to consider for the measurement update. First, consider the situation where the measurement pdf is in a normal, or Gaussian, form. Applying Baye's Theorem we can compute the updated pdf for our object:

$$\rho'(X) = \frac{\rho(X)\rho_H(X)}{\int_\infty \rho(X)\rho_H(X)dX}$$

Since the normal distribution has tails that extend to infinity, we find that the updated pdf is defined over the exact same phase volume set $\mathcal{B}_1$, and that the Gromov width is unchanged in this situation. The distribution of probability is redefined within this volume set, but the extent of the non-zero pdf and the bounding set $\mathcal{B}_1'$ remain unchanged. Repeated measurement can drastically change the distribution within this phase volume and its time mapping, but once a region has a finite probability a Gaussian measurement will never change its Gromov width. Furthermore, for a finite volume distribution the probability of finding the state within this distribution is always equal to unity. Thus, for a Gaussian measurement incorporated into a finite volume pdf both the total probability of the distribution is conserved as well as the Gromov width of that distribution. The only thing which changes is the redistribution of the pdf within these regions.

Now suppose we have a measurement pdf of a more general form, such that is also has a finite volume on which it is non-zero, characterized by the set in phase space $\mathcal{B}_z$. The resultant pdf ρ' will be defined over a non-zero probability set $\mathcal{B}_{z1} = \mathcal{B}_z \cap \mathcal{B}_1$. We find four different possibilities as a function of what set its intersection with $\mathcal{B}_1$ lies within.

1) If $\mathcal{B}_{z1} = \emptyset$, the empty set, the hypothesis that the measured object and the original object are consistent with each other is false.
2) If $\mathcal{B}_1 \subset \mathcal{B}_z$ then $\mathcal{B}_{z1} = \mathcal{B}_1$ and the symplectic width of the distribution is unchanged, although the distribution of probability does change, similar to the classical measurement case.
3) If $\mathcal{B}_z \subset \mathcal{B}_1$, then $\mathcal{B}_{z1} = \mathcal{B}_z$ and the Gromov width is completely determined by the measurement pdf and its associated volume $\mathcal{B}_z$ and $w_L(\mathcal{B}_{z1}) \leq w_L(\mathcal{B}_1)$.
4) If $\mathcal{B}_1 \not\subset \mathcal{B}_{z1}$ and $\mathcal{B}_z \not\subset \mathcal{B}_{z1}$, meaning that the two original distributions have regions where they do not overlap, the Gromov width decreases, $w_L(\mathcal{B}_{z1}) < \min\{w_L(\mathcal{B}_z), w_L(\mathcal{B}_1)\}$, and there is a change in the distribution probability.

We note that even for pdfs defined over a finite volume, the rules on the Gromov width change if instead of looking at volumes of unity probability one considers volumes of non-unity probability. Consider a set $\mathcal{B}_r$ such that $P(\mathcal{B}_r) = r < 1$ and perform a measurement update on the overall distribution (assumed to be a Gaussian measurement pdf for definiteness). While the given set $\mathcal{B}_r$ is unchanged, the probability density within it is changed and the new probability associated with this set is not conserved. If the probability is to be conserved, the Gromov width will decrease in general as it will correspond to a different, more concentrated set in phase volume. This was seen explicitly in the discussion on linear systems with Gaussian distributions.

VI. Conclusions

This paper describes the connection between uncertainty distributions, as represented by probability density functions, and symplectic capacities for Hamiltonian Dynamical Systems. We review some new derivations of the constraint on covariance matrices that arises from topological considerations, previously reported elsewhere. We also provide an explicit discussion of how the symplectic width of probability distributions becomes modified under the action of measurements and find situations where the symplectic width can be conserved under measurement.

References

[1] M. Gromov, Pseudoholomorphic curves in symplectic manifolds. Inventiones Mathematica 82 (1985) 307–347.

[2] I. Ekeland and H. Hofer, Symplectic topology and Hamiltonian dynamics, I and II. Mathematische Zeitschrift 200 (1990) 355–378 and 203 (1990) 553–567.

[3] D.J. Scheeres, F.-Y. Hsiao, R.S. Park, B.F. Villac, and J.M. Maruskin, Fundamental Limits on Spacecraft Orbit Uncertainty and Distribution Propagation, Journal of the Astronautical Sciences 54 (2006) 505–523.

[4] F.-Y. Hsiao and D.J. Scheeres, Fundamental Constraints on Uncertainty Relations in Hamiltonian Systems, IEEE Transactions In Automatic Control 52(4) (2007) 686–691.

[5] J.M. Maruskin, D.J. Scheeres, and A.M. Bloch, Dynamics of Symplectic Subvolumes, SIAM Journal of Applied Dynamical Systems 8(1) (2009) 180–2001.

[6] D.J. Scheeres, M.A. de Gosson, and J.M. Maruskin, Applications of Symplectic Topology to Orbit Uncertainty and Spacecraft Navigation, Journal of the Astronautical Sciences, in press.

[7] H.P. Robertson, The uncertainty principle. Physical Reviews 34 (1929) 163–164.

[8] E. Schrödinger, Zum Heisenbergschen Unschärfeprinzip, Berliner Berichte (1930), 296–303 [English translation: A. Angelow, M.C. Batoni, About Heisenberg Uncertainty Relation. Bulgarian Journal of Physics 26 5/6 (1999) 193–203, and http://arxiv.org/abs/quant-ph/9903100.

[9] A. Milani, G. Gronchi, M. Vitturi, and Z. Knezevic. 2004. "Orbit determination with very short arcs. i admissible regions." *Celestial Mechanics & Dynamical Astronomy* 90, 5785.

[10] R.S. Park and D.J. Scheeres. 2006. "Nonlinear Mapping Of Gaussian State Uncertainties: Theory And Applications To Spacecraft Control And Navigation," 2006. *Journal of Guidance, Control and Dynamics* 29(6): 1367-1375.

[11] H. Hofer and E. Zehnder, Symplectic Invariants and Hamiltonian Dynamics, Birkhäuser Advanced texts (Basler Lehrbücher), Birkhäuser Verlag (1994).

[12] J. Williamson, On the algebraic problem concerning the normal forms of linear dynamical systems. American Journal of Mathematics 58 (1936) 141–163.

[13] M. de Gosson, The Symplectic Camel and the Uncertainty Principle: The Tip of an Iceberg? Foundations of Physics 99 (2009) 194–214.

[14] L. Polterovich, The Geometry of the Group of Symplectic Diffeomorphisms, Lectures in Mathematics, Birkhäuser (2001).