The Discreteness-driven Relaxation of Collisionless Gravitating Systems: Entropy Evolution and the Nyquist–Shannon Theorem

Leandro Beraldo e Silva¹,², Walter de Siqueira Pedra³, and Monica Valluri¹

¹Universidade de São Paulo, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Departamento de Astronomia, CEP 05508-090, São Paulo, SP, Brasil
²Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA
³Universidade de São Paulo, Instituto de Física, Departamento de Física Matemática, CP 66318, CEP 05314-970, São Paulo, SP, Brasil

Abstract

The time irreversibility and fast relaxation of collapsing N-body gravitating systems (as opposed to the time reversibility of the equations of motion for individual stars or particles) are traditionally attributed to information loss due to coarse graining in the observation. We show that this subjective element is not necessary once one takes into consideration the fundamental fact that these systems are discrete, i.e., composed of a finite number, N, of stars or particles. We show that a connection can be made between entropy estimates for discrete systems and the Nyquist–Shannon sampling criterion. Specifically, given a sample with N points in a space of d dimensions, the Nyquist–Shannon criterion constrains the size of the smallest structures defined by a function in the continuum that can be uniquely associated with the discrete sample. When applied to an N-body system, this theorem sets a lower limit to the size of phase-space structures (in the continuum) that can be resolved in the discrete data. As a consequence, the finite N system tends to a uniform distribution after a relaxation time that typically scales as N₁/d. This provides an explanation for the fast achievement of a stationary state in collapsing N-body gravitating systems such as galaxies and star clusters, without the need to advocate for the subjective effect of coarse graining.

Key words: galaxies: evolution – galaxies: formation – galaxies: general – galaxies: kinematics and dynamics

1. Introduction

An important question in the study of collisionless N-body gravitating systems is how to reconcile the time irreversibility of the fast relaxation of a collapsing structure with the reversible character of the equations of motion for the individual stars or particles. In other words, how to reconcile this irreversible relaxation with the time reversibility of the Vlasov equation (assumed to describe the kinetic evolution). The most common solution to this apparent paradox (“the fundamental paradox of stellar dynamics,” according to Ogorodnikov 1965, p. 119) is to attribute this time irreversibility to information loss in a coarse-grained observation (Lynden-Bell 1967; Levin et al. 2014). According to this view, during the dynamical evolution of the system, the distribution of particles in phase space progressively develops finer and finer structures (i.e., filaments) that after some time can no longer be detected by the observing device, which only measures averaged (coarse-grained) quantities.

A fundamental problem with this solution, however, is that it introduces a subjective element, making the relaxation phenomenon dependent on the observational precision (Jaynes 1965). In this paper, we provide an alternative scenario without this subjective element. We use rigorous recipes to estimate the entropy of a discrete sample (see Joe 1989; Beirland et al. 1997; Leonenko et al. 2008; Biau & Devroye 2015) and show that its evolution is connected to the celebrated Nyquist–Shannon (N-S) sampling theorem of signal theory and image processing.

For systems with N particles evolving in a phase space of dimension d, we derive a relaxation time that scales typically as N₁/d. Once one recognizes as a fundamental fact that gravitational systems such as galaxies and star clusters are finite N systems, as opposed to the theoretical limit to the continuum (N → ∞), this timescale is naturally seen as a real relaxation time. Our analysis provides a theoretical explanation for the power-law N-dependencies of the relaxation time obtained by Pakter & Levin (2017) for long-range interacting systems in d = 2 and by Beraldo e Silva et al. (2019) for ensembles of orbits integrated in triaxial gravitational potentials (d = 6). Finally, together with the analysis of Beraldo e Silva et al. (2017, hereafter Paper I) and Beraldo e Silva et al. (2019, hereafter Paper II), this discreteness effect is shown to be the main mechanism for the fast (in a few dynamical timescales) collisionless relaxation of non-equilibrium N-body gravitating systems.

In Section 2 we show analytically that the finest phase-space structures of an ensemble of free particles are expected to develop linearly in time. In Section 3 we briefly introduce the N-S theorem, relating the inverse size of these finest phase-space structures (bandwidth) with the size N of a discrete sample. In Section 4 we develop a non-dynamical toy model with known values of the bandwidth, and in Section 5 we introduce the entropy estimator, showing that the estimates applied to the toy model agree with the N-S criterion. In Section 6 we apply the entropy estimator to the study of the relaxation of orbit ensembles integrated in a Plummer potential, deriving the typical relaxation time. Finally, Section 7 summarizes our results.

2. The Simplest Dynamical Model

In this section we consider the simple example of an ensemble of free particles to show how the phase-space structures are expected to evolve in time. In particular, in order to make contact with the N-S theorem in the next sections, we are interested in the time evolution of the finest phase-space structures, i.e., of the largest structures (the bandwidth K) in Fourier space of the distribution function.
Consider an ensemble of particles initially distributed according to a Gaussian in a \(d\)-dimensional phase space,

\[
f_0(x, v) = \frac{1}{(2\pi)^{d/2}} \exp \left( -\frac{x^2 + v^2}{2} \right),
\]

where \(x\) and \(v\) are conveniently normalized dimensionless position and velocity. If no forces act on the particles, at a time \(t\) this distribution evolves to

\[
f_t(x, v) = \frac{1}{(2\pi)^{d/2}} \exp \left( -\frac{(x - vt)^2 + v^2}{2} \right).
\]

Note that if Equation (2) is seen as characterizing the macroscopic state, i.e., considering the system as a whole, as is normally the case when referring to a (probability) distribution function, it associates a number to each point \((x, v)\) in a continuous \(d\)-dimensional domain for every time \(t\). In this way, when describing the evolution of a finite \(N\) system, it necessarily extrapolates the information available in a discrete sample, implicitly assuming that this extrapolation is always physically meaningful.

Taking the Fourier transform of Equation (2), one gets (see the Appendix for more general and detailed calculation)

\[
f_\hat{t}(\hat{x}, \hat{v}) = \frac{1}{(2\pi)^{d/2}} \int dx \int dv \, e^{-i (\hat{x} x + \hat{v} v)} \exp \left( \frac{-(x^2 + v^2)}{2} \right)
\]

\[
= \exp \left( \frac{-((\hat{x}^2 + 1)\hat{x}^2 + 2\hat{x} \cdot \hat{v} + \hat{v}^2)}{2} \right),
\]

where \((\hat{x}, \hat{v})\) are the respective wave vectors, or frequencies, in Fourier space. We are interested in determining the largest characteristic scales, i.e., the bandwidth \(K\), in Fourier space and a rough estimate can be made as

\[
K^2(t) \approx \langle \hat{x}^2 + \hat{v}^2 \rangle = \frac{\int d\hat{x} \int d\hat{v} \langle \hat{x}^2 + \hat{v}^2 \rangle f_\hat{t}(\hat{x}, \hat{v})}{\int d\hat{x} \int d\hat{v} f_\hat{t}(\hat{x}, \hat{v})} = (2 + t^2).
\]

This already shows that at large times the bandwidth \(K\) is expected to grow linearly with time for this simple model. A different estimate takes into account that the characteristic scales of \(f_\hat{t}(\hat{x}, \hat{v})\) can be associated with the dispersions in two orthogonal directions obtained by some rotation of the \((\hat{x}, \hat{v})\) axes. Since we are interested in the largest characteristic scales, we make a rotation such that one of the new axis points in the direction of the largest dispersion. This is similar to the so-called principal component analysis and is made in the Appendix, where we similarly conclude that \(K(t) \propto t\) for large times.

3. N-S Theorem

According to the N-S theorem, a function in the continuum can be recovered from its discrete sampling whenever the sampling rate is at least twice the bandwidth \(K\) in Fourier space of the function (N-S criterion). In the original proofs of Nyquist (1928) and Shannon (1949), the sampling was assumed to be uniform, but their result was extended, later on, to the case of nonuniform samplings, in which the sampling rate is to be understood as the average sampling rate (Landau 1967). This sufficient condition to recover a function in the continuum from some discrete sampling of it is also known to be necessary, in general. According to this theorem, in order to exactly reconstruct a function in a \(d\)-dimensional continuum (i.e., in a continuous domain) from a discrete sample, the number of sampling points must be \(N \geq K^n\), where \(K\) is the largest characteristic frequency of the function, i.e., its bandwidth in frequency/Fourier space. Conversely, given an arbitrary discrete sample with \(N\) points, the theorem states that only functions with a bandwidth

\[
K \lesssim N^{1/d},
\]

i.e., with structures not finer than given by Equation (3), can be uniquely associated to the sample. Functions with a larger bandwidth, i.e., with finer structures, contain extra information not equivalent to that in the sample.

4. Toy Model

In this section we introduce a (non-dynamical) toy model based on a simple probability distribution function for \(d\) independent variables

\[
f(x_1, \ldots, x_d) = F(x_1) \ldots F(x_d),
\]

with

\[
F(x) = \frac{1}{A} \left[ 1 + \sum_{m=1}^n \left[ a_m \cos(2\pi mk x) + b_m \sin(2\pi mk x) \right] \right]
\]

for \(-1/2 \leq x \leq 1/2\) and \(F(x) = 0\) otherwise. Here \(k, m, n\) are natural numbers and \(A\) is a normalization constant. \(A = 2\) if \(k = 0\) and \(A = 1\) when \(k > 0\). Note that the case \(k = 0\) corresponds to a uniform distribution. In order to ensure that \(F(x)\) is non-negative, the coefficients \(a_m, b_m\) are real numbers such that \(\sum_n \sqrt{a_n^2 + b_n^2} \leq 1\).

In the examples discussed below, for fixed \(n > 0\) we set \(b_m = 0\) for all \(m\) and \(a_m = n^{-1}\) for \(m \leq n\), with \(a_m = 0\) for \(m > n\). We denote by \(f^{(n)}\) the distributions corresponding to these choices of \(a_m, b_m\). There is nothing special about this choice, except that the coefficients vanish for \(m > n\). Other choices with this property lead to essentially the same results. For this model the bandwidth is \(K = n \times k\). Note that this toy model has no a priori dynamical interpretation or time evolution. However, the parameter \(k\) can be seen as analogous to time \(t\), with larger values introducing finer structures in the phase space and a linear growth of the bandwidth \(K\).

It is important to emphasize that, on the one hand, the imposition of a function in the continuum restricts the number of sampling points necessary to correctly recover the function. On the other hand, in \(N\)-body gravitating systems, what we are given a priori is an arbitrary sample of size \(N\). In this case, the N-S theorem imposes restrictions on the distribution function in the continuum that can be used to describe the \(N\)-body system and the equation driving its kinetic evolution. This point is discussed in Section 6.

Figure 1 shows samples of this model, Equations (4) and (5), with \(n = 1\) in \(d = 2\) with \(N = 10^4\) points generated with the acceptance-rejection method for different \(k\) values. For small \(K\) values, the structure of the distribution function is well captured by the sample. Larger values of \(K\) are associated with finer structures in the “phase space”.
when we calculate

\[ k p D d ^{-} = N k d \]

refers to the estimate of the quantity, and not to its Fourier transform. Different from Section 2 and the Appendix, in what follows, the symbol \( \hat{f} \) refers to the estimate of the quantity, and not to its Fourier transform.

Given a sample of points in \( d \) dimensions, distributed according to \( f \), this entropy can be estimated as

\[ S(k) \equiv - \int f \ln f \, dx_1 \ldots dx_d. \tag{6} \]

Given a sample of points in \( d \) dimensions, distributed according to \( f \), this entropy can be estimated as

\[ \hat{S}(k) = - \frac{1}{N} \sum_{i=1}^{N} \ln \hat{f}_i, \tag{7} \]

where the integral in Equation (6) is translated into a sum over the \( N \) sampling points. Equation (7) converges to Equation (6) for \( N \to \infty \) when we calculate \( \hat{f}_i \) with at least two methods (Joe 1989; Beirlant et al. 1997; Leonenko et al. 2008; Biau & Devroye 2015): the nearest neighbor method and the kernel method. A study of the evolution of \( N \)-body self-gravitating systems has shown that both methods provide very similar entropy estimates (see Paper I).

In the nearest neighbor method, the distribution function \( f \) at the point \( x_i = (x_1, \ldots, x_d) \) is estimated as the number of points inside a hyper-sphere of radius \( D_{in} \) (the distance from point \( i \) to its nearest neighbor \( n \)) divided by its volume. With normalization factors,

\[ \hat{f}_i = \frac{1}{(N - 1) e^\gamma V_d D_m^d} \tag{8} \]

(see Leonenko et al. 2008), where \( \gamma \approx 0.57722 \) is the Euler-Mascheroni constant, \( V_d = \pi^{d/2} / \Gamma(d/2 + 1) \), and \( D_m = (x_j - x_i)^2 \). For the identification of the nearest neighbors we use the approximate nearest neighbor (ANN) method (Arya et al. 1998), which is based on a kd-tree algorithm (Friedman et al. 1977). The algorithm allows one to optimize the search by approximating the nearest neighbor, but we use it without any approximation, identifying the exact nearest neighbor.

The entropy estimates of samples generated with Equations (4) and (5) for \( n = 1 \) and \( d = 2, 4, 6 \) are shown in Figure 2. The black points of the left panel (2D) contain the entropy values for the samples shown in Figure 1. Note that the theoretical entropy value obtained with Equations (4)–(6) is \( S_t(k) = -d \times 0.30727 \) for any integer \( k \), whereas \( S_0(k = 0) = 0 \). Here, \( S_t \) is the Shannon entropy of the distribution \( f^{(n)} \) defined above. Thus, if the sample is able to recover the full information of the function \( f \) at a fixed \( k \) value, we get \( \Delta S \equiv \hat{S}(k) - \hat{S}(k = 1) \approx 0 \), up to small statistical fluctuations. This is approximately the case for small \( K \), as can be seen in Figures 1 and 2. However, for any given number \( N \) of points, there is a critical \( K \) value beyond which the function in the continuum \( f \) has structures too fine to be resolved or, equivalently, a too large bandwidth in frequency/Fourier space, for all of the information contained in the function \( f \) to be recovered from the discrete sample, the distribution of which becomes effectively uniform. In this case, \( \hat{S}(k) \) strongly deviates from the entropy of the function in the

\footnote{Available at www.cs.umd.edu/~mount/ANN/. A slightly different version, allowing searches in parallel, was developed by Andreas Girgensohn and was kindly provided by David Mount.}
continuum, $S(k)$, achieving the maximum, 0, associated to a uniform distribution, thus implying $\Delta S = \tilde{S}(k) - \tilde{S}(k = 1) \simeq d \times 0.307$, shown as horizontal dashed lines in Figure 2. This entropy increase is the imprint of the N-S criterion, as quantitatively demonstrated below.

Let us mention that there is an essential difference between our entropy estimator $\tilde{S}$ and the case covered by the N-S theorem: whereas the latter refers to a discrete sample of the actual values of the distribution $f$, only estimates $\tilde{f}_i$, $i = 1, \ldots, N$, of $f$ are available for computing $\tilde{S}$. Hence, the feature of entropy estimators discussed in this paper is a (strong) analogy, a sort of stochastic instance of the original N-S theorem.

The data in Figure 2 can be described by the function

$$\Delta \tilde{S}(K) = \frac{A}{\pi/2 + \arctan(BC)} \times [\arctan(B(K - C)) + \arctan(BC)],$$

(9)

where $A$, $B$, and $C$ are free parameters representing, respectively, the maximum entropy increase, the slope of the rising part of the entropy production curve in Figure 2, and its delay (i.e., the $K$ value where the entropy starts to increase). The term $\arctan(BC)$ on the right-hand side ensures that $\Delta \tilde{S}(K = 0) = 0$ and the term $\pi/2 + \arctan(BC)$ in the denominator guarantees that $A = \Delta \tilde{S}(K \rightarrow \infty)$. This function (solid lines in Figure 2) provides reasonable fits for all values of $d$ and $N$.

We now define $K_{\Delta S/2}$, the value of $K$ at which the entropy production achieves half of its asymptotic value. Then, with the help of Equation (9) we obtain

$$K_{\Delta S/2} = \sqrt{B^{-2} + C^2}.$$  

(10)

This quantity represents the critical $K$ value beyond which the information regarding the continuous function is not adequately captured by the discrete sample.

Figure 3 shows this quantity, calculated with the values of $B$ and $C$ obtained in the previous fits for $n = 1, 2, 3$ and $d = 2, 4, 6, 8$. The lines are power-law fits to these points. Note that different values of parameter $n$ represent different models (see Equation (5)) with “phase-space” structures different from those of Figure 1. For all of these different models we obtain approximately

$$K_{\Delta S/2} \propto N^{1/d},$$

(11)

in agreement with the N-S criterion (Equation (3)). This shows that the estimates agree with the entropy of the corresponding distribution function $f$ in the continuum, if the assumed function generating the sample fulfills the N-S criterion. This suggests that the information contents in the sample and the whole function are approximately equivalent when the criterion is satisfied.

6. Relaxation of Gravitating Systems

Having shown that the entropy estimates agree with the N-S criterion, i.e., that they capture the information available from a discrete sample, we now move on to the study of the relaxation
process of finite $N$ gravitating systems. Using the Agama Library (Vasiliev 2019), we integrate ensembles of orbits in the Plummer potential

$$\phi(r) = -\frac{GM}{a} \frac{1}{\sqrt{1 + (r/a)^2}}.$$  

(12)

where $G$ is the gravitational constant, $a$ is a scale radius, and $M$ is the total mass. Initial conditions are generated with particles sampling a top hat, i.e., a uniform sphere (both in positions and velocities) of radius $a$ and maximum velocity $v_{\text{max}} = \sqrt{2}\phi(a)$. We set $GM = a = 1$ and integrate the ensembles for $\approx 300\tau_{\text{cr}}$, with the crossing time estimated as $\tau_{\text{cr}} = 2\pi\sqrt{(r^2)/a^3}$, where these quantities are calculated at $t = 0$. The entropy is estimated with Equations (7)–(8), where each of the six phase-space coordinates is normalized by its initial inter-percentile range containing 68% of the data around the median.

Figure 4 shows the entropy evolution for ensembles of various sizes $N$ (different colors). We note the resemblance of these data with that of the toy model (Figure 2). Replacing $K$ by $t/\tau_{\text{cr}}$, Equation (9) again provides reasonable fits (solid lines). In accordance with the second law of thermodynamics, the time evolution of the system (initially in a non-stationary state) is such that the entropy increases up to a maximum, where it stabilizes. In Paper II, we show that the entropy is conserved for self-consistent (i.e., stationary) samples, also in agreement with the second law of thermodynamics.

In the analysis above, we impose a distribution function in the continuum, i.e., Equations (4) and (5) for the toy model and the top hat initial condition in this section, and ask ourselves how many data points we need to recover the information contained in this function. From this point of view, the use of a finite $N$ limits the possibility of recovering information contained in fine structures and the entropy increase appears as a result of information loss (coarse graining) in the entropy estimation.

However, for real gravitational systems such as galaxies and star clusters, the situation is quite the opposite: what is given a priori, i.e., the real data, is a discrete sample of $N$ stars (or particles), and the question is if the fine structures developed by the assumed distribution function in the continuum represent real effects (supported by the discrete data) or rather spurious features introduced by the theoretical model. Given a discrete sample, the N-S theorem guarantees a one-to-one correspondence with functions in the continuum whose characteristic frequencies satisfy the N-S criterion (Equation (3)). For larger frequencies (finer structures), many functions in the continuum can be associated to the same sample and the choice of one specific function (with structures finer than allowed by the sample) constitutes an information input not contained in the sample itself.

Note that the very notion of convergence of a sequence of distribution functions developing rapidly varying structures (filaments) with the time evolution is an important conceptual point: such sequences cannot converge in the point-wise sense and the so-called weak convergence is a more natural notion in this situation, as pointed out by Mouhot & Villani (2011). This type of convergence means, roughly, that structures that get arbitrarily fine in the limit must be averaged out in order to obtain a well-defined limiting distribution. Our approach sheds light on this question, by providing a quantitative criterion to objectively evaluate the collapse of fine structures in distributions of particles of macroscopic systems with fixed (finite) $N$.

Once one recognizes that the real data is a finite $N$ sample (as opposed to the limit $N \to \infty$), one can safely consider the entropy evolution in Figure 4 as characterizing a real relaxation effect. Analogously to Equation (10), we define the typical time for this entropy increase as the time when it achieves half of its asymptotic value:

$$\frac{T_{\Delta S/2}}{\tau_{\text{cr}}} = \sqrt{B^{-2} + C^2}.$$  

(13)

This quantity, calculated with the fitting values of parameters $B$ and $C$, is shown in Figure 5 (points). This timescale is well fitted by a power law, which we write as

$$\frac{T_{\Delta S/2}}{\tau_{\text{cr}}} \propto N^{1.04d},$$  

(14)

where $d = 6$ is the dimension of the phase space.

The power law obtained for the Plummer potential (Figure 5) implies $\alpha \approx 0.91$. In Paper II, the integration of orbit ensembles in an integrable triaxial potential gives $0.98 \lesssim \alpha \lesssim 1.12$, depending on the initial conditions. Finally, the
relaxation times obtained by Pakter & Levin (2017) in $d = 2$
}imply $\alpha$ ranging from 1 for an integrable system to $\approx 0.3$ for a
 highly chaotic one. This weakening of the $N$-dependence of the
 relaxation time was interpreted by Pakter & Levin (2017) as a
 consequence of an enhancement in the efficiency of phase mixing in
 the presence of chaotic motion, and the same effect seems to be present
 in the results of Paper II, where we found $\alpha \approx 0.85$ for orbit ensembles
 in a non-integrable cuspy triaxial potential. In light of the results presented
 in this paper, these outcomes can be interpreted as a direct consequence
 of the N-S sampling criterion.

Comparison of Equations (11) and (14) indicates that the
 dynamical evolution of the system is such that the bandwidth of
 its distribution function grows with time $t$ as

$$K \propto \left(\frac{t}{\tau_{\text{rel}}}\right)^{1/\alpha}. \quad (15)$$

In particular, for $K$ growing linearly in time, as in the simple case
 of free particles discussed in Section 2, we have $\alpha = 1$. The
 results quoted above suggest that an approximately linear time
 dependence happens for the evolution in integrable potentials
 in general. In such potentials, the use of angle-time dependence
 happens for the evolution in integrable harmonic potential leads
 to periodic solutions. Suitable technical conditions. Note, in
 particular, that the harmonic potential leads to periodic (instead of linear) behavior
 of the bandwidth and one cannot expect the conjecture to hold
 true for every integrable potential.

Concluding, the power-law $N$-dependence for the relaxation
 time (Equation (14)) can be seen as a direct consequence of
 the N-S theorem. A linear time growth ($\alpha \approx 1$) of the bandwidth
 in frequency/Fourier space of the distribution function for
 integrable systems provides a relaxation time scaling as
 $\propto N^{1/4}$ for such systems. Moreover, the results of Paper II
 show that for a realistic cuspy gravitational potential hosting
 large fractions of chaotic orbits, the relaxation timescale does not
 seem to change dramatically ($\alpha \approx 0.85$ in that case),
 at least for the models considered in Paper II.

7. Summary

To summarize, our results show that the subjective element
 associated to the necessity of coarse graining in order to explain
 the fast relaxation of forming or perturbed N-body gravitating
 systems can be eliminated, via the N-S sampling criterion, if
 one recognizes as a fundamental fact that these are finite $N$
 systems whose evolution saturates their distribution in phase
 space at some time determined essentially by the dimension $d$
 and sample size $N$ (and also by the complexity of trajectories
 in phase space). Then, a posteriori one can look for the
 distribution function (and for the effective equation driving
 its evolution) compatible with the information contained in the
 sample at each time. In particular, our results suggest that
 the typical relaxation time of integrable systems in a phase space
 of dimension $d$ can be roughly estimated as $T/\tau_{\text{rel}} \propto N^{1/4}$, with
 the presence of chaotic orbits accelerating the growth of the
 bandwidth (i.e., the production of finer structures) but not
 dramatically changing the relaxation time $T$-dependence. Note
 that this timescale is small, in comparison to the two-body
 (collisional) relaxation time, which scales as $T_{\text{col}}/\tau_{\text{rel}} \propto N/\ln N$
, even for systems containing a small number of stars, such as open clusters ($N \lesssim 10^4$).

We regard the connection between the N-S criterion (with
 the recognition of the discreteness of gravitational systems) and
 the entropy evolution shown in this paper as a fundamental theoretical element if one wants to understand the fast
 collisionless relaxation of collapsing gravitational structures. Interestingly, these results seem to be in line with
 the “Indispensability of Atomism in Natural Science” supported
 by Boltzmann (1974).

We thank Jean-Bernard Bru for interesting discussions and
 hints. This work has made use of the computing facilities of the
 Laboratory of Astroinformatics (IAG/USP, NAT/Unicsul),
 whose purchase was made possible by the Brazilian agency
 FAPESP (2009/54006-4) and the INCT-A. L.B.S. is supported
 by FAPESP (2014/23751-4 and 2017-01421-0). W.S.P. is
 supported by CNPq (308337/2017-4). M.V. acknowledges
 support from HST-AR-13890.001, NSF award AST-1515001,
 NASA-ATP award NN15AK79G. This paper made use of the
 Agama (Vasiliev 2019), ANN (Arya et al. 1998), GSL
 (Galassi et al. 2018), matplotlib (Hunter 2007), numpy (Walt et
 al. 2011), and scipy (Jones et al. 2001).

Appendix

Let $d \in \mathbb{N}$ be the phase-space dimension (in a $3D$ space,
 $d = 6$). We define the probability distribution
 $f_0 : \mathbb{R}^{d/2} \times \mathbb{R}^{d/2} \rightarrow \mathbb{R}^+$ by

$$f_0(x, v) \equiv \frac{1}{(2\pi\sigma_x\sigma_v)^{d/2}} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{v^2}{2\sigma_v^2}\right).$$

where $x, v \in \mathbb{R}^{d/2}$ are the particles positions and velocities, respectively. If no forces act on the particles (zero potential) then, at any time $t > 0$, the distribution evolves to

$$f_t(x, v) = \frac{1}{(2\pi\sigma_x\sigma_v)^{d/2}} \exp\left[-\frac{(x - vt)^2}{2\sigma_x^2} - \frac{v^2}{2\sigma_v^2}\right]$$

Taking the Fourier transform of $f_t$ one gets

$$\hat{f}_t(\hat{x}, \hat{v}) \equiv \frac{1}{(2\pi\sigma_x\sigma_v)^{d/2}} \int dx \int dv \exp$$

$$\times \left[-\frac{(x - vt)^2}{2\sigma_x^2} - \frac{v^2}{2\sigma_v^2}\right] \exp\{-i(\hat{x} \cdot x + \hat{v} \cdot v)\}$$

$$= \exp\left[-\frac{\sigma_x^2\hat{x}^2 + \sigma_v^2(\hat{v} \cdot \hat{x})^2}{2}\right] = \exp\left(-\frac{\hat{x}^T\hat{A}\hat{x}}{2}\right),$$

where we defined $\hat{A} \equiv (\hat{x}, \hat{v}) \in \mathbb{R}^d$ and the Hermitian matrix

$$A \equiv \begin{pmatrix} \frac{t^2 + s^2}{2} & t \\ t & 1 \end{pmatrix}.$$
where \( s^2 \equiv \sigma_x^2 / \sigma_v^2 \). At this point, note that a rough estimate of the bandwidth in Fourier space can be obtained as

\[
K^2(t) \approx \langle \hat{s}^2 + \hat{v}^2 \rangle = \int d\xi d\hat{s} d\xi d\hat{v} \hat{s}^2(\hat{\xi}, \hat{\nu})
\]

This already shows that, for large times \( t \), the bandwidth \( K(t) \) is expected to grow linearly with \( t \). A more precise estimate of \( K(t) \) involves identifying the principal directions, i.e., a system of orthogonal axes obtained from \( (\hat{\xi}, \hat{\nu}) \) by a rotation such that one of the new axes points in the direction of largest dispersion. For this, we diagonalize the matrix \( A \), concluding that

\[
\frac{1}{2} \left[ (s^2 + 1) + t^2 \left( 1 \pm \sqrt{1 + \frac{2(s^2 + 1)}{t^2} + \frac{(s^2 - 1)^2}{t^4}} \right) \right]
\]

are two eigenvalues and respective orthogonal eigenvectors. Expressing the vector \( \hat{X}_T = (\hat{\xi}, \hat{\nu}) \) in the orthonormal basis associated to the two eigenvectors above (note that they are not normalized), we conclude that

\[
\frac{\sigma_x}{2} \hat{X}^T A \hat{X} = \frac{\hat{X}^+ (t, \hat{\xi}, \hat{\nu})^2}{2\hat{\sigma}_{\hat{\nu} \hat{\nu}}} + \frac{\hat{X}^- (t, \hat{\xi}, \hat{\nu})^2}{2\hat{\sigma}_{\hat{\nu} \hat{\nu}}},
\]

where

\[
\hat{X}^+ (t, \hat{\xi}, \hat{\nu}) \equiv \sqrt{\frac{1}{2t} \left[ (s^2 - 1) + t^2 \left( 1 \pm \sqrt{1 + \frac{2(s^2 + 1)}{t^2} + \frac{(s^2 - 1)^2}{t^4}} \right) \right]} \hat{\xi} + \hat{\nu}
\]

and

\[
\hat{X}^- (t, \hat{\xi}, \hat{\nu}) \equiv \sqrt{\frac{1}{2t} \left[ (s^2 - 1) + t^2 \left( 1 \pm \sqrt{1 + \frac{2(s^2 + 1)}{t^2} + \frac{(s^2 - 1)^2}{t^4}} \right) \right]} \hat{\xi} - \hat{\nu}.
\]

Note that \( \hat{X}^+ (t, \hat{\xi}, \hat{\nu}) \) and \( \hat{X}^- (t, \hat{\xi}, \hat{\nu}) \) are two vectors in \( \mathbb{R}^{d/2} \) such that

\[
\hat{X}^+ (t, \hat{\xi}, \hat{\nu})^2 + \hat{X}^- (t, \hat{\xi}, \hat{\nu})^2 = \hat{\xi}^2 + \hat{\nu}^2 = \hat{X}^2,
\]

and that they point in the principal directions such that \( \hat{\sigma}_{\hat{\nu} \hat{\nu}} > 0 \) are the corresponding variances. Observe also that

\[
\hat{\sigma}_{\hat{\xi} \hat{\nu}}, \hat{\sigma}_{\hat{\xi} \hat{\nu}} = 1.
\]

This identity is related to the fact that phase-space volume is preserved by dynamics (Liouville theorem) together with the Parseval identity for the Fourier transform. At large \( t > 0 \), by a Taylor expansion, one has

\[
t^2 \left( 1 \pm \frac{1 - 2s^2 + 1}{t^2} + \frac{(s^2 - 1)^2}{t^4} \right)
\]

\[
t^2 \pm \left( t^4 - 2s^2 + 1 - \frac{2s^2}{t^2} \right) + O(t^{-4})
\]

Thus, the bandwidth \( K \), i.e., the largest scales in Fourier space (smallest scales in real space) of the distribution function, estimated here as the largest among the dispersions in the directions \( \hat{\xi} \) and \( \hat{\nu} \), grows as \( \sigma_v^{-2} t \), for large times \( t > 0 \). Moreover, in this simple example, the velocity components alone are responsible for the growth of \( K \) (development of fine structures of \( f(\xi, v) \)) for large times.

### ORCID iDs

Leandro Beraldo e Silva [https://orcid.org/0000-0002-0740-1507](https://orcid.org/0000-0002-0740-1507)

Monica Valluri [https://orcid.org/0000-0002-6257-2341](https://orcid.org/0000-0002-6257-2341)

### References

Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R., & Wu, A. Y. 1998, *J. ACM*, 45, 391

Beirán, J., Dudewicz, E. J., Györfi, L., & Van Der Meulen, E. C. 1997, *Int. J. Math. Stat. Sci.*, 6, 17

Beraldo e Silva, L., de Siqueira Pedra, W., Sodré, L., Perico, E. L. D., & Lima, M. 2017, *ApJ*, 846, 125

Beraldo e Silva, L., de Siqueira Pedra, W., Valluri, M., & Sodré, L. 2019, *ApJ*, 870, 128

Biau, G., & Devroye, L. 2015, *Lectures on the Nearest Neighbor Method* (Berlin: Springer)

Boltzmann, L. 1974, in *Theoretical Physics and Philosophical Problems: Selected Writings*, ed. B. McGuinness (Berlin: Springer)

Friedman, J. H., Bentley, J. L., & Finkel, R. A. 1977, *ACM Trans. Math. Softw.*, 3, 209

Galassi, M., et al. 2018, *GNU Scientific Library, Reference Manual* (3rd ed.)

Hunter, J. D. 2007, *CSE*, 9, 90

Jaynes, E. T. 1965, *AmJPR*, 33, 391
Joe, H. 1989, AISM, 41, 683
Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open Source Scientific Tools for Python, http://www.scipy.org/
Landau, H. J. 1967, AcMa, 117, 37
Leonenko, N., Pronzato, L., & Savani, V. 2008, Tatra Mt. Math. Publ., 39, 265
Levin, Y., Pakter, R., Rizzato, F. B., Teles, T. N., & Benetti, F. P. C. 2014, PhR, 535, 1
Lynden-Bell, D. 1967, MNRAS, 136, 101

Mouhot, C., & Villani, C. 2011, AcMa, 207, 29
Nyquist, H. 1928, Trans. Am. Inst. Electr. Eng., 47, 617
Ogorodnikov, K. F. 1965, in Dynamics of Stellar Systems, ed. A. Beer (Oxford: Pergamon)
Pakter, R., & Levin, Y. 2017, JSMTE, 4, 044001
Shannon, C. E. 1949, PIRE, 37, 10
Vasiliev, E. 2019, MNRAS, 482, 1525
Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22