Maximum-entropy description of animal movement

Chris H. Fleming, Yiğit Subaşı, and Justin M. Calabrese

Phys. Rev. E **91**, 032107 — Published 4 March 2015

DOI: 10.1103/PhysRevE.91.032107

A maximum-entropy description of animal movement

Chris H. Fleming,[1, 2] Yiğit Subaşi,[3] and Justin M. Calabrese[1]
[1]*Smithsonian Conservation Biology Institute, Front Royal, Virginia*
[2]*Department of Biology, University of Maryland College Park*
[3]*Department of Chemistry and Biochemistry, University of Maryland College Park*

(Dated: January 30, 2015)

We introduce a class of maximum-entropy states that naturally includes within it all of the major continuous-time stochastic processes that have been applied to animal movement, including Brownian motion, Ornstein–Uhlenbeck motion, integrated Ornstein–Uhlenbeck motion, a recently discovered hybrid of the previous models, and a new model that describes central-place foraging. We are also able to predict a further hierarchy of new models that will emerge as data quality improves to better resolve the underlying continuity of animal movement. Finally, we also show that Langevin equations must obey a fluctuation-dissipation theorem to generate processes that fall from this class of maximum-entropy distributions when the constraints are purely kinematic.

I. INTRODUCTION

Animal movement is a continuous-time process, exhibiting continuous velocities and accelerations. Animal locations are, however, measured coarsely in time, even by modern GPS technology. The degree of continuity that we can actually resolve is limited by our measurement apparatus. To confront this scenario, we derive a natural class of maximum-entropy states for stochastic processes that are assumed to be very continuous, but are only sampled at discrete time intervals. The constraints with which we maximize entropy equate to understanding that a finite sampling frequency can only resolve the continuity of the sampled process to a finite degree. As for all other behaviors of the process, we are guided by the principle of maximum entropy.

The class of maximum-entropy states we derive is found to include within it Brownian motion (BM) [1–3], Ornstein–Uhlenbeck (OU) motion [4–7], integrated OU motion [8–10], and a more general movement model that includes all of the previous models as limiting cases [11]. In contrast with Brownian motion, which diffuses endlessly, the OU process is bound to a finite domain. Therefore, OU motion was originally applied to animal-tracking data to estimate the home-range areas of range-resident species [5]. On the other hand, as neither Brownian nor OU motion provide a differentiable process, integrated OU (IOU) motion was introduced to estimate instantaneous velocities in noisy telemetry datasets [8]. Brownian motion has a long history of use in describing animal movement and, despite its aforementioned limitations, the Brownian-motion movement model has still found use in estimating distributions of occurrence from tracking data with large gaps between observations [2, 3]. Brownian motion is a special case of both OU motion, in the limit of an infinite home-range area, and IOU motion, in the limit of impersistent movement. Recently, animal tracking datasets that are sampled finely enough to estimate velocity and long enough to estimate home-range behavior have spurred the development of an OUF model that generalizes OU and IOU motion [11, 12]. Here, we provide a theoretical framework that explains this coincident grouping of movement models in terms of continuity and entropy, and predicts a missing model within the same group that corresponds to central-place foraging. We can also predict what models will become appropriate as GPS and battery technology improve to the point that more of the underlying continuity of animal movement is revealed.

Finally, we find that the multidimensional generalizations of these stochastic models obey a fluctuation-dissipation theorem (FDT) when the maximum-entropy constraints are purely kinematic, in that they do not impose dynamical relationships between the process and its derivatives. On the other hand, if dynamical constraints are allowed, then the resulting class of maximum-entropy states is larger than what can be generated by ordinary Langevin equations without a FDT.

In thermodynamic systems, fluctuations and dissipation are engendered by the same microscopic degrees of freedom, even though they are phenomenologically distinct. As a simple example, for a damped mechanical system with position $\mathbf{x}(t)$ driven by thermal white noise $\boldsymbol{\xi}(t)$, the Langevin equation is given by [13]

$$\mathbf{M}\,\ddot{\mathbf{x}}(t) + \underbrace{2\,\mathbf{\Gamma}\,\dot{\mathbf{x}}(t)}_{\text{dissipation}} - \mathbf{F}(\mathbf{x}(t)) = \underbrace{\boldsymbol{\xi}(t)}_{\text{fluctuations}} \quad , \tag{I.1}$$

where $\mathbf{M}$ is the mass, $\mathbf{F}(\mathbf{x})$ is the net deterministic force on the system, and $\mathbf{\Gamma}$ is the motion damping coefficient, which drives relaxation to the rest state. In our analogous biological equations of motion, the average trajectory follows a deterministic mean $\boldsymbol{\mu}(t)$, which can be thought of as being generated by deterministic "forces", while random movements about the mean manifest from fluctuations, and relaxation back to the mean arises from dissipation. In classical non-equilibrium thermodynamics, the fluctuations and dissipation are related by [14]

$$\langle \boldsymbol{\xi}(t)\,\boldsymbol{\xi}(t')^{\mathrm{T}} \rangle = \tilde{\boldsymbol{\sigma}}_{\xi\xi}\,\delta(t-t')\,, \qquad \tilde{\boldsymbol{\sigma}}_{\xi\xi} = 2 k_B T\,\mathbf{\Gamma}\,, \tag{I.2}$$

here for white noise, where $\tilde{\boldsymbol{\sigma}}_{\xi\xi}$ is the spectral density of the fluctuations, k_B is Boltzmann's constant and T is the temperature of the surrounding environment. More generally, yet still in the classical regime, dissipation can be nonlocal in time and fluctuations can be colored, but a proportionality relation still holds between the dissipation kernel and noise correlation [14]. In quantum non-equilibrium thermodynamics, proportionality between the dissipation kernel and noise correlation still holds in the frequency domain, though the proportionality constant is nonlinear in T [15, 16]. Finally, even when the environment is not in equilibrium and cannot be assigned a temperature, the dissipation kernel and noise correlation must always satisfy an inequality due to the Heisenberg uncertainty principle [17].

The thermodynamic FDT is necessary for microscopic theories of stochastic processes to be consistent with macroscopic thermodynamics. But we might imagine that dissipation coefficients and fluctuation autocorrelations are more generally unrelated—particularly in systems that have nothing to do with thermodynamics. For the maximum-entropy distributions we explore here, we find that the fluctuations and dissipation are not necessarily proportional, but they must obey non-trivial commutation relations.

II. MAXIMUM-ENTROPY STATES

In this multivariate treatment, we denote vectors and matrices in bold, such as the position $\mathbf{x}(t)$ and autocorrelation function $\boldsymbol{\sigma}(t,t')$ (defined in Eq. (A.4)). To constrain the degree of continuity in the underlying process, we will use the relationship between the continuity of the stochastic process and the continuity of its autocorrelation function. As we lack information about the higher-order cumulants of the process, we leave everything beyond the second cumulant to ignorance, and so upon constraining the autocorrelation function $\boldsymbol{\sigma}(t,t')$, the entropy per unit time functional is given by (App. A)

$$h[\boldsymbol{\sigma}] = \frac{1}{2}\int\!df\,\mathrm{tr}\log\tilde{\boldsymbol{\sigma}}(f) + \text{constant}\,, \tag{II.1}$$

in terms of the spectral-density function $\tilde{\boldsymbol{\sigma}}(f)$, defined by

$$\boldsymbol{\sigma}(t,t') = \int\!df\,e^{+2\pi\imath f(t-t')}\,\tilde{\boldsymbol{\sigma}}(f)\,, \tag{II.2}$$

for stationary autocorrelations, where $\boldsymbol{\sigma}(t,t') = \boldsymbol{\sigma}(t-t')$. We consider only stationary autocorrelations, because their diagonalizing basis elements $e^{+2\pi\imath ft}$ are universally defined and stationary autocorrelations can be considered as the time average of non-stationary autocorrelations when estimating their parameters from a non-stationary process [12]. Importantly, we do not assume that the mean function (A.3) is stationary. Furthermore, as we show in App. A, the entropy functional (II.1) still holds for non-stationary autocorrelations, though under a different transformation basis. Therefore our derived maximum-entropy states also hold for non-stationary autocorrelations, though under a different set of constraints.

A. Variance constraint

As a simple example, we first consider a process with only its variance constrained to $\boldsymbol{\sigma}(0)$ and no further information:

$$\boldsymbol{\sigma}(\tau)\big|_{\tau=0} = \boldsymbol{\sigma}(0)\,, \qquad \int\!df\,\tilde{\boldsymbol{\sigma}}(f) = \boldsymbol{\sigma}(0)\,, \tag{II.3}$$

where the latter relation is conveniently expressed in the frequency domain. The quantity to maximize, with Lagrange multiplier $\boldsymbol{\lambda}_0/2$, is given by

$$m[\boldsymbol{\sigma}] = h[\boldsymbol{\sigma}] + \frac{1}{2}\mathrm{tr}\,\boldsymbol{\lambda}_0\left(\boldsymbol{\sigma}_0 - \int\!df\,\tilde{\boldsymbol{\sigma}}(f)\right). \tag{II.4}$$

Using matrix derivatives [18, App. B], the Euler-Lagrange equations are then given by

$$\frac{1}{2}\tilde{\boldsymbol{\sigma}}(f)^{-\mathrm{T}} = \frac{1}{2}\boldsymbol{\lambda}_0^{\mathrm{T}}\,, \qquad \tilde{\boldsymbol{\sigma}}(f) = \boldsymbol{\lambda}_0^{-1}\,, \tag{II.5}$$

which implies that the spectral-density function is a constant matrix. I.e., the maximum-entropy process with variance $\boldsymbol{\sigma}(0)$ is a white-noise process with variance $\boldsymbol{\sigma}(0)$. The maximum-entropy process is not correlated in time without providing any further kinematic constraints.

B. Kinematic constraints and continuity

The k^{th} derivative of $\mathbf{x}(t)$ has the autocorrelation function

$$\frac{d^k}{dt^k}\frac{d^k}{dt'^k}\left\langle[\mathbf{x}(t)-\boldsymbol{\mu}(t)]\,[\mathbf{x}(t')-\boldsymbol{\mu}(t')]^{\mathrm{T}}\right\rangle = \frac{d^k}{dt^k}\frac{d^k}{dt'^k}\boldsymbol{\sigma}(t,t')\,. \tag{II.6}$$

Placing a constraint upon the k^{th} derivative of $\mathbf{x}(t)$ to have variance $\boldsymbol{\sigma}^{(k)}(0)$ takes the form

$$(-1)^k\frac{\partial^{2k}}{\partial\tau^{2k}}\boldsymbol{\sigma}(\tau)\bigg|_{\tau=0} = \boldsymbol{\sigma}^{(k)}(0)\,, \tag{II.7}$$

$$\int\!df\,(2\pi f)^{2k}\,\tilde{\boldsymbol{\sigma}}(f) = \boldsymbol{\sigma}^{(k)}(0)\,. \tag{II.8}$$

After maximizing entropy with these constraints, the spectral-density function is then given by

$$\tilde{\boldsymbol{\sigma}}(f) = \left[\sum_{k=0}^{K}(2\pi f)^{2k}\boldsymbol{\lambda}_{2k}\right]^{-1}, \tag{II.9}$$

when including kinematic constraints up to order K.

As any differentiable function is continuous, if a process has derivatives that always take finite values, then this process is always continuous. Therefore, by placing kinematic constraints up to order K, we ensure that the process is continuous with $K-1$ continuous derivatives. The K^{th} derivative of the process is not continuous, but is a well defined white-noise process.

An important property of the spectral-density function (II.9) is that it is an even function of the frequency f, which implies that the autocorrelation function $\boldsymbol{\sigma}(\tau)$ is a symmetric function of the time lag τ. More generally, cross correlations between different spatial dimensions or different individual animals can be asymmetric in time, which allows for leader-follower or time-lagged relationships. This can occur when dynamical constraints are imposed in addition to the kinematic constraints (App. B).

1. *K=1: OU & BM motion*

As we have already shown, $K = 0$ corresponds to uncorrelated motion of a particular variance. $K = 1$ corresponds to Ornstein–Uhlenbeck motion [5–7], which is a continuous process with autocorrelation function

$$\sigma(\tau) = \sigma(0)\, e^{-f|\tau|}\,, \tag{II.10}$$

in one dimension. This model describes Brownian motion bound to a finite area around the mean location [4], and the timescale $1/f$ determines how long it takes the animal to cross its home-range area.

The semi-variance function [11] is a standard measure of diffusion [19] and for stationary processes it is given by

$$\gamma(\tau) \equiv \frac{1}{2}\left\langle \left|x(t{+}\tau) - \mu(t{+}\tau) - x(t) + \mu(t)\right|^2 \right\rangle, \tag{II.11}$$

$$= \sigma(0) - \sigma(\tau)\,, \tag{II.12}$$

with the latter relation holding for processes with a finite variance. Ordinary Brownian motion is a limiting case of OU motion for small f, whereupon the semi-variance function limits to

$$\lim_{f\to0}\gamma(\tau) = D\,|\tau|\,, \qquad D = \lim_{f\to0}\sigma(0)\,f\,, \tag{II.13}$$

where D is the diffusion rate. This linear dependence upon lag τ by $\gamma(\tau)$ is referred to as regular diffusion. For the OU process, the diffusion is regular for small lags, as can be seen from a Taylor-series expansion of $\gamma(\tau)$, while for larger lags the diffusion is limited by the existence of a finite variance $\sigma(0)$.

2. *K=2: OUF & IOU motion*

$K = 2$ includes within it OUF motion, which extends the OU model to continuous velocities, thought possibly to explain the ballistic foraging behavior of Mongolian gazelles [11]. The OUF autocorrelation function is given by

$$\sigma(\tau) = \sigma(0)\,\frac{f_+\,e^{-f_-|\tau|} - f_-\,e^{-f_+|\tau|}}{f_+ - f_-}\,, \qquad f_- < f_+\,, \tag{II.14}$$

This model describes correlated velocity movement within a home range of variance $\sigma(0)$. At timescales finer than $1/f_+$, motion appears ballistic as

$$\gamma(\tau) = \frac{1}{2}\sigma_v(0)\,\tau^2 + \mathcal{O}(\tau^3)\,, \qquad \sigma_v(0) = f_+ f_-\,\sigma(0)\,, \tag{II.15}$$

while at timescales coarser than $1/f_+$, motion is OU-like with a home-range crossing time of $1/f_-$.

In the limit of uncorrelated velocities, OUF motion reduces to OU motion as

$$\lim_{f_+\to\infty}\sigma(\tau) = \sigma(0)\,e^{-f_-|\tau|}\,. \tag{II.16}$$

On the other hand, in the limit of unbounded motion, OUF motion simplifies to integrated OU motion [8–10], as the semi-variance function limits to

$$\lim_{f_-\to0}\gamma(\tau) = \frac{\sigma_v(0)}{f_+^2}\left(f_+|\tau| - 1 + e^{-f_+|\tau|}\right), \tag{II.17}$$

which can be differentiated twice to obtain the OU velocity autocorrelation function

$$\lim_{f_-\to0}\sigma_v(\tau) = \sigma_v(\tau)\,e^{-f_+|\tau|}\,. \tag{II.18}$$

Like Brownian motion, the integrated OU process exhibits unbounded diffusion. However, IOU velocities only vary by a finite amount $\sigma_v(0)$ from the mean $\dot{\mu}(t)$, and relax back the mean with characteristic rate f_+. IOU motion is ballistic at timescales finer than $1/f_+$. Some peculiarities of purely ballistic diffusion are discussed in [20].

3. *K=2: Central-place foraging*

Considering the general structure of Eq. (II.9), there is one remaining model included in $K = 2$ that has not previously been considered in the movement-ecology literature:

$$\sigma(\tau) = \sigma(0)\,e^{-f|\tau|}\left(\cos\omega\tau + \frac{f}{\omega}\sin\omega|\tau|\right). \tag{II.19}$$

In this model there are periodic episodes of diffusion from and relaxation back to the mean location μ. The phenomenological behavior of this model is particularly relevant for describing central-place foraging [21], where an animal has a nest or den at its mean location μ and periodically leaves to perform a random search for resources. This periodic motion stands in contrast to periodicities in the mean, such as migration, where the animal cycles between its summering and wintering grounds. The probability density of a central-place forager is unimodal, whereas the probability density of a migratory species is bi-modal.

4. K=1: An excluded model

It is also interesting to note what models are not included in this class. For instance, the autocorrelation function

$$\sigma(\tau) = \sigma(0)\, e^{-f|\tau|} \cos(\omega\tau)\,, \tag{II.20}$$

does not have a spectral-density function consistent with Eq2. (II.9) or (B.1) with any finite number of constraints, even though this model can be considered as an oscillatory generalization of the OU process in other, physical contexts [22–24].

C. Multi-variate Ornstein–Uhlenbeck motion

Constraining the process up to its velocity results in the spectral-density function

$$\tilde{\boldsymbol{\sigma}}(f) = \left[\boldsymbol{\lambda}_0 + (2\pi f)^2 \boldsymbol{\lambda}_2\right]^{-1}, \tag{II.21}$$

where both $\boldsymbol{\lambda}$ matrices must be positive definite for this to be a valid spectral-density function. Factoring this expression, we have

$$\tilde{\boldsymbol{\sigma}}(f) = \boldsymbol{\lambda}_2^{-\frac{1}{2}} \left[\mathbf{F}^2 + (2\pi f)^2\right]^{-1} \boldsymbol{\lambda}_2^{-\frac{1}{2}}\,, \tag{II.22}$$

where $\mathbf{F}^2 = \boldsymbol{\lambda}_2^{-1/2} \boldsymbol{\lambda}_0\, \boldsymbol{\lambda}_2^{-1/2}$ must then be a positive-definite matrix of square frequencies. Fourier transforming back into the time domain, we have the autocorrelation function

$$\boldsymbol{\sigma}(\tau) = \boldsymbol{\lambda}_2^{-\frac{1}{2}} \frac{e^{-|\tau|\mathbf{F}}}{2\,\mathbf{F}} \boldsymbol{\lambda}_2^{-\frac{1}{2}}\,. \tag{II.23}$$

This describes a multivariate Ornstein–Uhlenbeck process of various dissipation rates and unit variance $\mathbf{I}$ that is linearly transformed to have variance

$$\boldsymbol{\sigma}(0) = \frac{1}{2}\, \boldsymbol{\lambda}_2^{-\frac{1}{2}}\, \mathbf{F}^{-1}\, \boldsymbol{\lambda}_2^{-\frac{1}{2}}\,. \tag{II.24}$$

III. FLUCTUATION-DISSIPATION THEOREM

A. Ornstein–Uhlenbeck theorem

To compare with Eq. (II.23) and without loss of generality, we will consider the Langevin equation of a multivariate, mean-zero OU process $\mathbf{x}(t)$, which represents the difference between the animal's location and its mean:

$$\dot{\mathbf{x}}(t) = \underbrace{-\boldsymbol{\Gamma}\,\mathbf{x}(t)}_{\text{dissipation}} + \underbrace{\boldsymbol{\xi}(t)}_{\text{fluctuations}}, \quad \langle \boldsymbol{\xi}(t)\, \boldsymbol{\xi}(t')^{\mathrm{T}} \rangle = \tilde{\boldsymbol{\sigma}}_{\xi\xi}\, \delta(t-t')\,. \tag{III.1}$$

where $\boldsymbol{\xi}(t)$ is a zero-mean white-noise process. In terms of animal movement, the "fluctuation" terms in Langevin equation (III.8) represent the random aspect of movement that gives rise to diffusion, while the "dissipation" terms generate advection back towards the mean location and keep the animal bound within a well defined home-range area despite its tendency to move about randomly.

Note that the spectral density $\tilde{\boldsymbol{\sigma}}_{\xi\xi}$ must be positive definite and real, and therefore it is symmetric. Standardizing our Langevin equation so that the fluctuations have unit spectral density, we have

$$\dot{\mathbf{y}}(t) = -\mathbf{G}\, \mathbf{y}(t) + \mathbf{u}(t)\,, \tag{III.2}$$

in terms of the transformed variables

$$\mathbf{y}(t) = \tilde{\boldsymbol{\sigma}}_{\xi\xi}^{-\frac{1}{2}}\, \mathbf{x}(t)\,, \qquad \mathbf{u}(t) = \tilde{\boldsymbol{\sigma}}_{\xi\xi}^{-\frac{1}{2}}\, \boldsymbol{\xi}(t)\,, \tag{III.3}$$

$$\mathbf{G} = \tilde{\boldsymbol{\sigma}}_{\xi\xi}^{-\frac{1}{2}}\, \boldsymbol{\Gamma}\, \tilde{\boldsymbol{\sigma}}_{\xi\xi}^{+\frac{1}{2}}\,. \tag{III.4}$$

The dissipation matrices $\boldsymbol{\Gamma}$ and $\mathbf{G}$ are related by a similarity transform and therefore they share the same eigenvalues, but in general they will not share the same symmetries. Transforming to the frequency domain, we have

$$2\pi\imath f\, \tilde{\mathbf{y}}(f) = -\mathbf{G}\, \tilde{\mathbf{y}}(f) + \tilde{\mathbf{u}}(f)\,, \tag{III.5}$$

$$\tilde{\mathbf{y}}(f) = \left[2\pi\imath f + \mathbf{G}\right]^{-1} \tilde{\mathbf{u}}(f)\,, \tag{III.6}$$

and with this the spectral-density function is given by

$$\tilde{\boldsymbol{\sigma}}(f) = \langle \tilde{\mathbf{x}}(f)\, \tilde{\mathbf{x}}(f)^\dagger \rangle = \tilde{\boldsymbol{\sigma}}_{\xi\xi}^{+\frac{1}{2}} \langle \tilde{\mathbf{y}}(f)\, \tilde{\mathbf{y}}(f)^\dagger \rangle\, \tilde{\boldsymbol{\sigma}}_{\xi\xi}^{+\frac{1}{2}}\,, \tag{III.7}$$

$$= \tilde{\boldsymbol{\sigma}}_{\xi\xi}^{+\frac{1}{2}} \left[\mathbf{G} + 2\pi\imath f\right]^{-1} \left[\mathbf{G} + 2\pi\imath f\right]^{-\dagger} \tilde{\boldsymbol{\sigma}}_{\xi\xi}^{+\frac{1}{2}}\,. \tag{III.8}$$

From Eq. (II.22), if this is to represent a maximum-entropy state, then we must have

$$\tilde{\boldsymbol{\sigma}}_{\xi\xi} = \boldsymbol{\lambda}_2^{-1}\,, \qquad \mathbf{G}\,\mathbf{G}^{\mathrm{T}} = \mathbf{F}^2\,, \qquad \mathbf{G} = \mathbf{G}^{\mathrm{T}}\,. \tag{III.9}$$

This final symmetry, applied to Eq. (III.4), implies that the dissipation matrix and autocorrelation matrix must commute in the sense of

$$\left[\boldsymbol{\Gamma}, \tilde{\boldsymbol{\sigma}}_{\xi\xi}\right]_{\mathrm{T}} = \boldsymbol{\Gamma}\, \tilde{\boldsymbol{\sigma}}_{\xi\xi} - \tilde{\boldsymbol{\sigma}}_{\xi\xi}\, \boldsymbol{\Gamma}^{\mathrm{T}} = \mathbf{0}\,, \tag{III.10}$$

which reduces to ordinary commutation if $\boldsymbol{\Gamma}$ is symmetric. We refer to this relation as comprising the Ornstein–Uhlenbeck fluctuation-dissipation theorem. This FDT is more general (and weaker) than the thermodynamic relation, where the two matrices are strictly proportional.

B. General theorem

The analogous Langevin equation for a continuous process $\mathbf{x}(t)$ with mean zero and $K-1$ continuous derivatives is given by

$$\left[\frac{d}{dt}\right]^K \mathbf{x}(t) + \sum_{k=1}^{K} \boldsymbol{\Gamma}_k \left[\frac{d}{dt}\right]^{K-k} \mathbf{x}(t) = \boldsymbol{\xi}(t)\,. \tag{III.11}$$

By a similar procedure we have the transformed solutions

$$\tilde{\boldsymbol{y}}(f) = \left[(2\imath\pi f)^K + \sum_{k=1}^{K}(2\imath\pi f)^{K-k}\mathbf{G}_k\right]^{-1} \tilde{\boldsymbol{u}}(f) \,, \tag{III.12}$$

and for the spectral-density function to take the form (II.9), we must have the transformed commutation relations

$$\mathbf{G}_1 = \mathbf{G}_1^{\mathrm{T}} \,, \qquad \mathbf{G}_k^{\mathrm{T}}\,\mathbf{G}_{k+1} = \mathbf{G}_{k+1}^{\mathrm{T}}\,\mathbf{G}_k \,, \tag{III.13}$$

which then implies the commutation relations

$$\boldsymbol{\Gamma}_1\,\tilde{\boldsymbol{\sigma}}_{\xi\xi} = \tilde{\boldsymbol{\sigma}}_{\xi\xi}\,\boldsymbol{\Gamma}_1^{\mathrm{T}} \,, \quad \boldsymbol{\Gamma}_k\,\tilde{\boldsymbol{\sigma}}_{\xi\xi}^{-1}\,\boldsymbol{\Gamma}_{k+1}^{\mathrm{T}} = \boldsymbol{\Gamma}_{k+1}\,\tilde{\boldsymbol{\sigma}}_{\xi\xi}^{-1}\,\boldsymbol{\Gamma}_k^{\mathrm{T}} \,. \tag{III.14}$$

IV. RANGE-RESIDENCE VERSUS CENTRAL-PLACE FORAGING

In one dimension the mean-zero Langevin equation for $K = 2$ is given by

$$\ddot{x}(t) + 2\,f\,\dot{x}(t) + F^2\,x(t) = \xi(t) \,, \tag{IV.1}$$

which is the equation of motion of a simple, damped harmonic oscillator driven by white noise. Central-place foraging corresponds to the under-damped regime with relaxation rate f and foraging frequency ω parameters

$$f^2 < F^2 \,, \qquad \omega = \sqrt{F^2 - f^2} \,, \tag{IV.2}$$

where ω determines the frequency with which foraging bouts occur and f determines the rate at which correlations decay between successive foraging bouts. In central-location foraging, the animal periodically leaves its mean location to search for resource patches and returns. Just as a thermodynamic environment sets the Lagrange multiplier T to its temperature, an animal's environment can determine the animal's foraging frequency ω, which is often fixed to $2\pi/\mathrm{day}$.

For the range-resident OUF model, which corresponds to the over-damped regime, the two relaxation rates are given by

$$f^2 > F^2 \,, \qquad f_\pm = f \pm \sqrt{f^2 - F^2}, \tag{IV.3}$$

where the smaller f_- roughly determines the amount of correlation in successive positions and the larger f_+ roughly determines the amount of correlation in successive velocities. In range-resident motion, the animal exhibits autocorrelated velocities within a finite home range. Specifically for Mongolian gazelles, it has been observed that f_- is associated with the seasonal timescale [25], and so this Lagrange multiplier is also likely set by the environment.

V. DISCUSSION

This framework takes all of the major continuous-time, animal-movement models that have been used for empirical reasons—because they have appropriate phenomenology—and puts them into a coherent theoretical framework. The principle of maximum entropy provides a useful perspective as our constraints are very natural for animal location data, in that, animal movement is extremely continuous, yet location data are relatively coarse. This leads to a hierarchy of models whereupon an increasing degree of continuity can be modeled and all finer scale behaviors are conceded to ignorance. As GPS and battery technology improve, or possibly by combining current relocation and accelerometery data, our theory predicts that we can increase the number of kinematic constraints K to derive more suitable models.

A fluctuation-dissipation theorem is derived, which provides important criteria for model construction in cases where there is significant spatial anisotropy in the movement dynamics or interactions between individuals that diffuse at different rates. There can be significant variability in movement parameters among the individuals of population [12], and so if individual animals coordinate their movements then they can exhibit a mixture of movement modes. Our FDT constrains these interactions, which would arise from the off-diagonal terms of Eq. (III.8), and reduces the number of multivariate models that would need to be considered for model selection. Importantly, an information criteria can be used to compare the constrained, parameter-sparse and unconstrained, parameter-rich models and determine if our FDT finds empirical support. Therefore, these two findings—the maximum-entropy models and their FDT—provide a concrete example of how ideas and techniques from physics can contribute to other disciplines.

From a biological perspective, it is interesting that range-residence and central-place foraging can be considered as two parameter regimes of the same model, because they correspond to different movement strategies that are typically found in different taxonomic groups. Central-place foragers, like many birds and canids, maintain a nest or den at their mean location, while range-resident species, like moose and gazelle, are only found in the vicinity of their mean location, which is more of a statistical abstraction. Even though the major timescales of this theory are likely set by the environment, it does not seem reasonable to imagine that all parameters can be adjusted, for a given species, such that a transition between movement behaviors occurs. This fact suggests the hypothesis that the remaining timescales might be set by the evolutionary lineage that the species descends from, which can still be considered as part of the "environment" with respect to the system-environment partitioning of the physics formalism.

There are some mathematical similarities between our class of maximum-entropy states Burg's for discrete-time processes [26]. By constraining the autocorrelation function up to a fixed number of lags, Burg derived the entire class of discrete-time auto-regressive (AR) processes. Our constraints are better suited to our system of interest, and while we derive a class of continuous-time auto-regressive processes, they are restricted to obey a fluctuation-dissipation theorem when only kinematic constraints are considered. On the other hand, if dynamical constraints are considered (App. B), states are obtained that do not correspond to continuous-time auto-regressive processes.

A natural question that arises from this perspective regards how strong the analogy between our maximum-entropy states and thermodynamics might be. In both cases, the entropy is maximized with respect to natural constraints that regard what we can reasonably measure; in both cases there are Lagrange multipliers that are determined by the environment; and finally, in both cases there is a fluctuation-dissipation theorem, though in our case it is comparatively weak. In Brownian motion derived from Hamiltonian mechanics, there will always be a relationship between the fluctuations and dissipation, even outside of the context of thermodynamics [17], and so we might ask if there is any unifying microscopic theory that generates the FDT here and what sort of interpretation it has.

Appendix A: Derivation of entropy functional

Here we will show a result that is well known for multivariate random variables—if we constrain ourselves to the first two cumulants or moments of the stochastic process, then the distribution that maximizes entropy is the normal distribution. Alternatively, the mean function can be left unconstrained and the same results can be derived. The entropy of a distribution p is given by

$$H[p] = -\int \mathcal{D}\mathbf{x}\, p[\mathbf{x}] \log p[\mathbf{x}] \,, \tag{A.1}$$

and we will maximize it under the constraints

$$\begin{aligned} 1 &= \int \mathcal{D}\mathbf{x}\, p[\mathbf{x}] \,, && \text{(A.2)} \\ \boldsymbol{\mu}(t) &= \int \mathcal{D}\mathbf{x}\, p[\mathbf{x}]\, \mathbf{x}(t) \,, && \text{(A.3)} \\ \boldsymbol{\sigma}(t,t') &= \int \mathcal{D}\mathbf{x}\, p[\mathbf{x}]\, \left[\mathbf{x}(t) - \boldsymbol{\mu}(t)\right] \left[\mathbf{x}(t') - \boldsymbol{\mu}(t')\right]^{\mathrm{T}} \,, && \text{(A.4)} \end{aligned}$$

which is equivalent to maximizing

$$\begin{aligned} M[p] = H[p] &+ \lambda_0 \left(1 - \int \mathcal{D}\mathbf{x}\, p[\mathbf{x}]\right) + \int dt\, \boldsymbol{\lambda}_1(t)^{\mathrm{T}} \left(\boldsymbol{\mu}(t) - \int \mathcal{D}\mathbf{x}\, p[\mathbf{x}]\, \mathbf{x}(t)\right) \\ &+ \iint dt\, dt' \,\mathrm{tr}\, \boldsymbol{\lambda}_2(t,t') \left(\boldsymbol{\sigma}(t,t') - \int \mathcal{D}\mathbf{x}\, p[\mathbf{x}]\, \left[\mathbf{x}(t) - \boldsymbol{\mu}(t)\right] \left[\mathbf{x}(t') - \boldsymbol{\mu}(t')\right]^{\mathrm{T}}\right) . \end{aligned} \tag{A.5}$$

where the λ are Lagrange multipliers. The Euler-Lagrange equations are then given by

$$\begin{aligned} \log p[\mathbf{x}] + 1 &= \lambda_0 + \int dt\, \boldsymbol{\lambda}_1(t)^{\mathrm{T}} \mathbf{x}(t) + \iint dt\, dt' \,\mathrm{tr}\, \boldsymbol{\lambda}_2(t,t') \left[\mathbf{x}(t) - \boldsymbol{\mu}(t)\right] \left[\mathbf{x}(t') - \boldsymbol{\mu}(t')\right]^{\mathrm{T}} \,, && \text{(A.6)} \\ p[\mathbf{x}] &= \exp\!\left(\lambda_0 - 1 + \int dt\, \boldsymbol{\lambda}_1(t)^{\mathrm{T}} \mathbf{x}(t) + \iint dt\, dt' \,\mathrm{tr}\, \boldsymbol{\lambda}_2(t,t') \left[\mathbf{x}(t) - \boldsymbol{\mu}(t)\right] \left[\mathbf{x}(t') - \boldsymbol{\mu}(t')\right]^{\mathrm{T}}\right) . && \text{(A.7)} \end{aligned}$$

Choosing the Lagrange multipliers that satisfy our constraints, we finally have

$$p[\mathbf{x}] = \frac{1}{\sqrt{\det 2\pi\boldsymbol{\sigma}}} e^{-\frac{1}{2}\iint dt\, dt' \left[\mathbf{x}(t') - \boldsymbol{\mu}(t')\right]^{\mathrm{T}} \boldsymbol{\sigma}^{-1}(t,t') \left[\mathbf{x}(t') - \boldsymbol{\mu}(t')\right]} \,. \tag{A.8}$$

which is the distribution of a Gaussian stochastic process. The entropy of a Gaussian stochastic process is then given by

$$\begin{aligned} H[\boldsymbol{\mu},\boldsymbol{\sigma}] =& \frac{1}{2}\int\!\mathcal{D}\mathbf{x}\, p[\mathbf{x}] \left(\log\det 2\pi\boldsymbol{\sigma} + \iint\! dt\, dt' \,[\mathbf{x}(t)-\boldsymbol{\mu}(t)]^{\mathrm{T}}\, \boldsymbol{\sigma}^{-1}(t,t')\, [\mathbf{x}(t')-\boldsymbol{\mu}(t')]\right), \\ =& \frac{1}{2}\left(\int\! dt\, \mathrm{tr}\,(\log\boldsymbol{\sigma})(t,t) + \iint\! dt\, dt' \mathrm{tr}\,\mathbf{I}\,\delta(t-t')\right) + \text{constant}\,, && \text{(A.9)} \\ =& \frac{1}{2}\int\! dt\, \mathrm{tr}\,(\log\boldsymbol{\sigma})(t,t) + \text{constant}\,, && \text{(A.10)} \end{aligned}$$

which only depends on the autocorrelation function and not the mean. The mean is deterministic and does not carry with it any entropy.

Viewing the autocorrelation function $\boldsymbol{\sigma}(t,t')$ as a large, positive-definite matrix, its eigen-decomposition is given by

$$\boldsymbol{\sigma}(t,t') = \int\! df\, \mathbf{U}(t,f)\, \tilde{\boldsymbol{\sigma}}(f)\, \mathbf{U}(t',f)^{\dagger}\,, \tag{A.11}$$

$$\delta(t-t')\,\mathbf{I} = \int\! df\, \mathbf{U}(t,f)\, \mathbf{U}(t',f)^{\dagger}\,, \tag{A.12}$$

$$\delta(f-f')\,\mathbf{I} = \int\! dt\, \mathbf{U}(t,f)^{\dagger}\, \mathbf{U}(t,f')\,, \tag{A.13}$$

where for stationary autocorrelations $\mathbf{U}(f,t)$ is a harmonic function, $\tilde{\boldsymbol{\sigma}}(f)$ is the spectral-density function, and f is their frequency. The entropy functional is then given by

$$\begin{aligned} H[\boldsymbol{\sigma}] &= \frac{1}{2}\int\! dt \int\! df\, \mathrm{tr}\big[\mathbf{U}(t,f)\, \log\tilde{\boldsymbol{\sigma}}(f)\, \mathbf{U}(t,f)^{\dagger}\big]\,, && \text{(A.14)} \\ &= \frac{1}{2}\int\! df\, \mathrm{tr}\log\tilde{\boldsymbol{\sigma}}(f) \int\! dt\, \mathbf{U}(t,f)^{\dagger}\mathbf{U}(t,f)\,, && \text{(A.15)} \\ &= \frac{\delta(0)}{2}\int\! df\, \mathrm{tr}\log\tilde{\boldsymbol{\sigma}}(f)\,, && \text{(A.16)} \end{aligned}$$

to within a constant. The entropy functional is, in general, an infinite quantity. Equivalently, we may consider the the average entropy per unit time

$$h[\boldsymbol{\sigma}] = \frac{1}{2}\int\! df\, \mathrm{tr}\log\tilde{\boldsymbol{\sigma}}(f) + \text{constant}\,, \tag{A.17}$$

which is a finite quantity. For stationary autocorrelation functions, this functional is not only the average entropy rate but also the instantaneous entropy rate.

Appendix B: Dynamical Constraints

Placing dynamical constraints on the system, such as a cross correlation between $\mathbf{x}(t)$ and $\dot{\mathbf{x}}(t)$, allows for odd powers of frequency in the spectral-density function. To produce a valid autocorrelation function we must have an Hermitian spectral-density function, which requires that the more general form

$$\tilde{\boldsymbol{\sigma}}(f) = \left[\sum_{k=0}^{K}(2\pi f)^{2k}\boldsymbol{\lambda}_{2k} + \sum_{j=0}^{K-1}(2\pi\imath f)^{2j+1}\boldsymbol{\Lambda}_{2j+1}\right]^{-1}, \tag{B.1}$$

have anti-Hermitian Lagrange multipliers $\boldsymbol{\Lambda}_j = -\boldsymbol{\Lambda}_j^{\dagger}$ in addition to the Hermitian Lagrange multipliers $\boldsymbol{\lambda}_k = \boldsymbol{\lambda}_k^{\dagger}$. With this structure of autocorrelation function, there are maximum-entropy states that cannot be generated by an ordinary Langevin equation, even without a FDT constraining its dynamics.

As an example, for the generalized multivariate Ornstein–Uhlenbeck process, where the cross-correlation between position and velocity $\langle[\mathbf{x}(t)-\boldsymbol{\mu}(t)]\,[\dot{\mathbf{x}}(t')-\dot{\boldsymbol{\mu}}(t')]^{\mathrm{T}}\rangle$ is constrained in addition to the covariances of position and velocity, the spectral-density function is given by

$$\begin{aligned} \tilde{\boldsymbol{\sigma}}(f) &= \big[\boldsymbol{\lambda}_0 + (2\pi\imath f)\boldsymbol{\Lambda}_1 + (2\pi f)^2\boldsymbol{\lambda}_2\big]^{-1}\,, && \text{(B.2)} \\ &= \boldsymbol{\lambda}_2^{-\frac{1}{2}}\, \big[\mathbf{F}_2^2 + (2\pi\imath f)\mathbf{F}_1 + (2\pi f)^2\big]^{-1}\, \boldsymbol{\lambda}_2^{-\frac{1}{2}}\,, && \text{(B.3)} \end{aligned}$$

but if this spectral-density function is to be generated by an ordinary Langevin equation with time-local dissipation and white noise, then by Eq. (III.8) it must be equivalent to

$$\tilde{\boldsymbol{\sigma}}(f) = \tilde{\boldsymbol{\sigma}}_{\xi\xi}^{+\frac{1}{2}}\, \big[\mathbf{G}^{\mathrm{T}}\mathbf{G} + (2\pi\imath f)(\mathbf{G}^{\mathrm{T}}-\mathbf{G}) + (2\pi f)^2\big]^{-1}\, \tilde{\boldsymbol{\sigma}}_{\xi\xi}^{+\frac{1}{2}}\,. \tag{B.4}$$

Furthermore, the Lagrange multipliers must relate to the dynamical parameters via

$$\boldsymbol{\lambda}_2^{-1} = \tilde{\boldsymbol{\sigma}}_{\xi\xi}\,, \quad \mathbf{F}_2^2 = \mathbf{G}\,\mathbf{G}^{\mathrm{T}}\,, \quad \mathbf{F}_1 = \mathbf{G}^{\mathrm{T}} - \mathbf{G}\,, \tag{B.5}$$

which is not general enough to produce all valid maximum-entropy states. To see this, one can start with Eq. (B.4) and add a positive-definite matrix to $\mathbf{G}\,\mathbf{G}^{\mathrm{T}}$ in $\mathbf{F}_2^2$, which results in a positive-definite spectral-density function that cannot be factored into a single Hermitian square.

[1] A. Einstein, Annalen der Physik **322**, 549 (1905).

[2] J. S. Horne, E. O. Garton, S. M. Krone, and J. S. Lewis,

Ecology **88**, 2354 (2007).

[3] B. Kranstauber, R. Kays, S. D. LaPoint, M. Wikelski, and K. Safi, Journal of Animal Ecology **81**, 738 (2012).

[4] G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930).

[5] J. E. Dunn and P. S. Gipson, Biometrics **33**, 85 (1977).

[6] D. R. Brillinger and B. S. Stewart, Canadian Journal of Statistics **26**, 431 (1998).

[7] P. G. Blackwell, Biometrika **90**, 613 (2003).

[8] D. S. Johnson, J. M. London, M.-A. Lea, and J. W. Durban, Ecology **89**, 1208 (2008).

[9] E. Gurarie and O. Ovaskainen, The American Naturalist **178**, 113 (2011).

[10] E. M. Hanks, M. B. Hooten, D. S. Johnson, and J. T. Sterling, PLoS ONE **6**, e22795 (2011).

[11] C. H. Fleming, J. M. Calabrese, T. Mueller, K. A. Olson, P. Leimgruber, and W. F. Fagan, The American Naturalist **183**, E154 (2014).

[12] C. H. Fleming, J. M. Calabrese, T. Mueller, K. A. Olson, P. Leimgruber, and W. F. Fagan, Methods in Ecology and Evolution **5**, 462 (2014).

[13] A. O. Caldeira and A. J. Leggett, Physica A **121**, 587 (1983).

[14] R. Kubo, Reports on Progress in Physics **29**, 255 (1966).

[15] H. Callen and T. Welton, Phys. Rev. **83**, 34 (1951).

[16] K. Lindenberg and B. J. West, Phys. Rev. A **30**, 568 (1984).

[17] C. H. Fleming, B. L. Hu, and A. Roura, Physical Review E **88**, 012102 (2013).

[18] Y. Subaşi, C. H. Fleming, J. M. Taylor, and B. L. Hu, Phys. Rev. E **86**, 061132 (2012).

[19] N. Cressie, *Statistics for spatial data*, revised ed. (Wiley, New York, 1993).

[20] L. C. Lapas, I. V. L. Costa, M. H. Vainstein, and F. A. Oliveira, Europhysics Letters **77**, 37004 (2007).

[21] G. H. Orians and N. E. Pearson, "On the theory of central place foraging," in *Analysis of ecological systems* (Ohio State University Press, Columbus, 1979) pp. 155–177.

[22] X. Zhao, W. Shi, L.-A. Wu, and T. Yu, Phys. Rev. A **86**, 032116 (2012).

[23] C. J. Broadbent, J. Jing, T. Yu, and J. H. Eberly, Annals of Physics **327**, 1962 (2014).

[24] J. Xu, X. Zhao, J. Jing, L.-A. Wu, and T. Yu, (2014), arXiv:1407.4178 [quant-ph].

[25] C. H. Fleming, T. Mueller, K. A. Olson, P. Leimgruber, W. F. Fagan, and J. M. Calabrese, (2014), *in review*.

[26] J. P. Burg, Ph.D. thesis, Stanford University (1975).