MDPI

Article

Entropic Dynamics on Gibbs Statistical Manifolds

Pedro Pessoa *[ID]**, Felipe Xavier Costa * and Ariel Caticha ***

Department of Physics, University at Albany (SUNY), Albany, NY 12222, USA
* Correspondence: ppessoa@albany.edu (P.P.); fxaviercosta@albany.edu (F.X.C.); acaticha@albany.edu (A.C)

Abstract: Entropic dynamics is a framework in which the laws of dynamics are derived as an application of entropic methods of inference. Its successes include the derivation of quantum mechanics and quantum field theory from probabilistic principles. Here, we develop the entropic dynamics of a system, the state of which is described by a probability distribution. Thus, the dynamics unfolds on a statistical manifold that is automatically endowed by a metric structure provided by information geometry. The curvature of the manifold has a significant influence. We focus our dynamics on the statistical manifold of Gibbs distributions (also known as canonical distributions or the exponential family). The model includes an "entropic" notion of time that is tailored to the system under study; the system is its own clock. As one might expect that entropic time is intrinsically directional; there is a natural arrow of time that is led by entropic considerations. As illustrative examples, we discuss dynamics on a space of Gaussians and the discrete three-state system.

1. Introduction

The original method of Maximum Entropy (MaxEnt) is usually associated with the names of Shannon [1] and Jaynes [2–5], although its roots can be traced to Gibbs [6]. The method was designed to assign probabilities on the basis of partial information in the form of expected value constraints and the central quantity, called entropy, which was interpreted as a measure of uncertainty or as an amount of missing information. In a series of developments starting with Shore and Johnson [7], with further contributions from other authors [8–12], the range of applicability of the method was significantly extended. In its new incarnation, the purpose of the method of Maximum Entropy, which will be referred as ME to distinguish it from the older version, is to update the probabilities from arbitrary priors when new information in the form of constraints is considered [13]. Highlights of the new method include: (1) A unified treatment of Bayesian and entropic methods which demonstrates their mutual consistency. (2) A new concept of entropy as a tool for reasoning that requires no interpretation in terms of heat, multiplicities, disorder, uncertainty, or amount of information. Indeed, entropy in ME needs no interpretation; it is a tool designed to perform a certain function—to update probabilities to accommodate new information. (3) A Bayesian concept of information defined in terms of its effects on the beliefs of rational agents—the constraints are the information. (4) The possibility of information that is not in the form of expected value constraints (we shall see an example below).

The old MaxEnt was sufficiently versatile for providing the foundations to equilibrium statistical mechanics [2] and to find application in a wide variety of fields such as economics [14], ecology [15,16], cellular biology [17,18], network science [19,20], and opinion dynamics [21,22]. As is the case with thermodynamics, all these applications are essentially static. MaxEnt has also been deployed to non-equilibrium statistical mechanics (see [23,24] and subsequent literature in maximum caliber, e.g., [25–27]) but the dynamics is not intrinsic to the probabilities; it is induced by the underlying Hamiltonian dynamics

of the molecules. For problems beyond physics there is a need for more general dynamical frameworks based on information theory.

The ME version of the maximum entropy method offers the possibility of developing a true dynamics of probabilities. It is a dynamics driven by entropy—an Entropic Dynamics (ED)—which is automatically consistent with the principles for updating probabilities. ED naturally leads to an "entropic" notion of time. Entropic time is a device designed to keep track of the accumulation of changes. Its construction involves three ingredients: one must introduce the notion of an instant, verify that these instants are suitably ordered, and finally one must define a convenient notion of duration or interval between successive instants. A welcome feature is that entropic time is tailored to the system under study; the system is its own clock. Another welcome feature is that such an entropic time is intrinsically directional—an arrow of time is generated automatically.

ED has been successful in reconstructing dynamical models in physics such as quantum mechanics [28,29], quantum field theory [30], and the renormalization group [31]. Beyond physics, it has been recently applied to the rhw fields of finance [32,33] and neural networks [34]. Here, we aim for a different class of applications of ED: to describe the dynamics of Gibbs distributions, also known as canonical distribution (exponential family) in statistical physics (statistics), since they are the distributions that are defined by a set of expected values constraint, namely sufficient statistics. Unlike the other cited papers on ED, here we will not focus on what the distributions are meant to represent. Other assumptions that would be specific to the modeled system are beyond the scope of the present article.

The goal is to study the ED that is generated by transitions from one distribution to another. The main assumptions are that changes happen and that they are not discontinuous. We do not explain why changes happen—this is a mechanics without a mechanism. Our goal is to venture an educated estimate of what changes one expects to happen. The second assumption is that systems evolve along continuous trajectories in the space of probability distributions. It also implies that the study of motion involves two tasks. The first is to describe how a single infinitesimal step occurs. The second requires a scheme for keeping track of how a large number of these short steps accumulate to produce a finite motion. It is the latter task that involves the introduction of the concept of time.

The fact that the space of macrostates is a statistical manifold—each point in the space is a probability distribution—has a profound effect on the dynamics. The reason is that statistical manifolds are naturally endowed with a Riemannian metric structure that is given by the Fisher–Rao information metric (FRIM) [35,36]; this structure is known as information geometry [37–39]. The particular case of Gibbs distributions leads to additional interesting geometrical properties (see e.g., [40,41]), which have been explored in the extensive work relating statistical mechanics to information geometry [42–49]. Information geometry has also been used as a fundamental concept for complexity measures [50–52].

In this paper, we tackle the more formal aspects of an ED on Gibbs manifolds and offer a couple of illustrative examples. The formalism is applied to two important sets of probability distributions: the space of Gaussians and the space of distributions for a three-state system, both of which can be written in the exponential form. Because these distributions are both well-studied and scientifically relevant, they can give us a good insight into how the dynamics work.

It is important to emphasize that Gibbs distributions are not restricted to the description of a system in thermal equilibrium. While it is true that, if one chooses the conserved quantities in Hamiltonian motion as the sufficient statistics, the resultant Gibbs distributions are the ones that are associated to equilibrium statistical mechanics, the Gibbs distribution can be defined for arbitrary choices of sufficient statistics, and the modeling endeavour includes identifying the ones that are relevant to the problem at hand. On the same note, the dynamics developed here are not a form of nonequilibrium statistical mechanics, which is driven by a underlying physical molecular dynamics, while the ED is completely agnostic of any microstate dynamics.

The article is organized, as follows: the next section discusses the space of Gibbs distributions and its geometric properties; Section 3 considers the ideas of ED; Section 4 tackles the difficulties associated with formulating ED on the curved space of probability distributions; Section 5 introduces the notion of entropic time; Section 6 describes the evolution of the system in the form of a differential equation; in Section 7, we offer two illustrative examples of ED on a Gaussian manifold and on a two-simplex.

2. The Statistical Manifold of Gibbs Distributions

2.1. Gibbs Distributions

The canonical or Gibbs probability distributions are the macrostates of a system. They describe a state of uncertainty regarding the microstate $x \in \mathcal{X}$ of the macroscopic system. A canonical distribution $\rho(x)$ is assigned by maximizing the entropy

$$S[\rho|q] = -\int dx\, \rho(x) \log \frac{\rho(x)}{q(x)} \tag{1}$$

relative to the prior $q(x)$ subject to n expected value constraints

$$\int dx\, \rho(x) a^i(x) = A^i, \quad \text{with} \quad i = 1 \ldots n, \tag{2}$$

and the normalization of $\rho(x)$. Typically, the prior $q(x)$ is chosen to be a uniform distribution over the space $\mathcal{X}$ so that it is maximally non-informative, but this is not strictly necessary. The n constraints, on the other hand, reflect the information that happens to be relevant to the problem. The resulting canonical distribution is

$$\rho(x|\lambda) = \frac{q(x)}{Z(\lambda)} \exp\left[-\lambda_i a^i(x)\right], \tag{3}$$

where $\lambda = \{\lambda_1 \ldots \lambda_n\}$ are the Lagrange multipliers that are associated to the expected value constraints, and we adopt the Einstein summation convention. The normalization constant is

$$Z(\lambda) = \int dx\, q(x) \exp\left[-\lambda_i a^i(x)\right] = e^{-F(\lambda)}, \tag{4}$$

where $F = -\log Z$ plays a role analogous to the free energy. The Lagrange multipliers $\lambda_i(A)$ are implicitly defined by

$$\frac{\partial F}{\partial \lambda_i} = A^i. \tag{5}$$

Evaluating the entropy (1) at its maximum yields

$$S(A) = -\int dx\, \rho(x|\lambda(A)) \log \frac{\rho(x|\lambda(A))}{q(x)} = \lambda_i(A) A^i - F(\lambda(A)). \tag{6}$$

which we shall call the macrostate entropy or (when there is no risk of confusion) just the entropy. Equation (6) shows that $S(A)$ is the Legendre transform of $F(\lambda)$: a small change dA^i in the constraints shows that $S(A)$ is indeed a function of the expected values A^i,

$$dS = \lambda_i dA^i \quad \text{so that} \quad \lambda_i = \frac{\partial S}{\partial A^i}. \tag{7}$$

One might think that defining dynamics on the family of canonical distributions might be too restricted to be of interest; however, this family has widespread applicability. Here, it has been derived using the method of maximum entropy, but historically it has also been known as the exponential family, namely the only family of distributions that possesses sufficient statistics. Interestingly, this was a problem that was proposed by Fisher [53] in the primordium of statistics and later proved independently by Pitman [54], Darmois [55], and Koopman [56]. The sufficient statistics turn out to be the functions $a_i(x)$ in (1). In Table 1,

we give a short list of the priors $q(x)$ and the functions $a_i(x)$ that lead to well-known distributions [41,57].

Table 1. Identification of sufficient statistics, priors and Lagrange multipliers for some well-known probability distributions.

Distribution	λ parameter	Suff. Stat.	Prior
Exponent Polynomial $\rho(x\|\beta) = \frac{\sqrt[k]{\beta}}{\Gamma(1+1/\beta)} e^{-\beta x^k}$	$\lambda = \beta$	$a(x) = x^k$	uniform
Gaussian $\rho(x\|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	$\lambda = \left(-\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2}\right)$	$a(x) = (x, x^2)$	uniform
Multinomial (k) $\rho(x\|\theta) = \frac{n!}{x_1!...x_k!}\theta_1^{x_1}\ldots\theta_k^{x_k}$	$\lambda = -\log(\theta_1, \theta_2, \ldots, \theta_k)$	$a = (x_1, \ldots x_k)$	$q(x) = \prod_{i=1}^{k} x_i!$
Poisson $\rho(x\|m) = \frac{m^x}{x!}e^{-m}$	$\lambda = -\log m$	$a(x) = x$	$q(x) = 1/x!$
Mixed power laws $\rho(x\|\alpha,\beta) = \frac{x^{-\alpha}e^{-\beta x}}{\beta^{\alpha-1}\Gamma(1-\alpha)}$	$\lambda = (\alpha, \beta)$	$a = (\log x, x)$	uniform

Naturally, the method of maximum entropy assumes that the various constraints are compatible with each other, so that the set of multipliers λ exists. It is further assumed that the constraints reflect physically relevant information, so that the various functions, such as $A^i(\lambda) = \frac{\partial}{\partial \lambda_i} F$ and $\lambda_i(A) = \frac{\partial}{\partial A^i} S$, which appear in the formalism, are both invertible and differentiable, and so that the space of Gibbs distributions is indeed a manifold. However, the manifold may include singularities of various kinds that are of particular interest, as they may describe phenomena, such as phase transitions [42,46].

2.2. Information Geometry

We offer a brief review of well known results concerning the information geometry of Gibbs distributions in order to establish the notation and recall some results that will be needed in later sections [38,40].

To each set of expected values $A = \{A^1, A^2, \ldots, A^n\}$, or to the associated set of Lagrange multipliers $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, there corresponds a canonical distribution. Therefore the set of distributions $\rho(x|\lambda)$ or, equivalently, $\rho(x|A)$ is a statistical manifold in which each point can be labelled by the coordinates λ or by A. Whether we choose λ or A as coordinates is purely a matter of convenience. The change of coordinates is implemented using

$$\frac{\partial A^i}{\partial \lambda_k} = -\frac{\partial^2 \log Z}{\partial \lambda_k \partial \lambda_i} = A^k A^i - \langle a^k a^i \rangle , \tag{8}$$

where we recognize the covariance tensor,

$$C^{ij} = \langle (a^i - A^i)(a^j - A^j) \rangle = -\frac{\partial A^i}{\partial \lambda_j} . \tag{9}$$

Its inverse is given by

$$C_{jk} = -\frac{\partial \lambda_j}{\partial A^k} = -\frac{\partial^2 S}{\partial A^j \partial A^k} , \tag{10}$$

that means the inverse covariant matrix C_{ij} is the Hessian of negative entropy in (6). This implies

$$C^{ij}C_{jk} = \frac{\partial A^i}{\partial A^k} = \delta^i_k . \tag{11}$$

Statistical manifolds are endowed with an essentially unique quantity to measure the extent to which two neighboring distributions $\rho(x|A)$ and $\rho(x|A + dA)$ can be distinguished from each other. This measure of distinguishability provides a statistical notion of distance, which is given by FRIM, $d\ell^2 = g_{ij} dA^i dA^j$ where

$$g_{ij} = \int dx \, \rho(x|A) \frac{\partial \log \rho(x|A)}{\partial A^i} \frac{\partial \log \rho(x|A)}{\partial A^j}. \tag{12}$$

For a broader discussion on the existence, derivation, and consistency of this metric, as well as its properties, see [38–40]. Here, it suffices to say that FRIM is the unique metric structure that is invariant under Markov embeddings [58,59] and, therefore, is the only way of assigning a differential geometry structure that is in accordance to the grouping property of probability distributions.

To calculate g_{ij} for canonical distributions, we use

$$g_{ij} = \frac{\partial \lambda_k}{\partial A^i} \frac{\partial \lambda_l}{\partial A^j} \int dx \, \rho \, \frac{\partial \log \rho}{\partial \lambda_k} \frac{\partial \log \rho}{\partial \lambda_l} \tag{13}$$

and

$$\frac{\partial \log \rho(x|A)}{\partial \lambda_k} = A^k - a^k(x) \tag{14}$$

so that, using (8)–(11), we have

$$g_{ij} = C_{ik} C_{lj} C^{kl} = C_{ij} . \tag{15}$$

Therefore, the metric tensor g_{ij} is the inverse of the covariance matrix C^{ij}, which, by (10), is the Hessian of the entropy.

As mentioned above, instead of A^i, we could use the Lagrange multipliers λ_i as coordinates. Subsequently, the information metric is the covariance matrix,

$$g^{ij} = \int dx \, \rho(x|\lambda) \frac{\partial \log \rho(x|\lambda)}{\partial \lambda_i} \frac{\partial \log \rho(x|\lambda)}{\partial \lambda_j} = C^{ij} . \tag{16}$$

Therefore, the distance $d\ell$ between neighboring distributions can be written in either of two equivalent forms,

$$d\ell^2 = g_{ij} dA^i dA^j = g^{ij} d\lambda_i d\lambda_j . \tag{17}$$

Incidentally, the availability of a unique measure of volume $dV = (\det g_{ij})^{1/2} d^n A$ implies that there is a uniquely defined notion of the uniform distribution over the space of macrostates. The uniform distribution P_u assigns equal probabilities to equal volumes, so that

$$P_u(A) d^n A \propto g^{1/2} d^n A \quad \text{where} \quad g = \det g_{ij} . \tag{18}$$

To conclude this overview section, we note that the metric tensor g_{ij} can be used to lower the contravariant indices of a vector to produce its dual covector. Using (10) and (12), the covector dA_i dual to the infinitesimal vector with components dA^i is

$$dA_i = g_{ij} dA^j = -\frac{\partial \lambda_i}{\partial A^j} dA^j = -d\lambda_i . \tag{19}$$

which shows that not only are the coordinates A and λ related through a Legendre transformation, which is a consequence of entropy maximization, but also through a vector-covector duality, i.e., $-d\lambda_i$ is the covector dual to dA^i, which is a consequence of information geometry.

3. Entropic Dynamics

Having established the necessary background, we can now develop an entropic framework to describe the dynamics on the space of macrostates.

3.1. Change Happens

Our starting assumption is that changes happen continuously, which is supported by observation in nature. Therefore, the dynamics that we wish to formulate assumes that the system evolves along continuous paths. This assumption of continuity represents a significant simplification, because it implies that a finite motion can be analyzed as the accumulation of a large number of infinitesimally short steps. Thus, our first goal will be to find the probability $P(A'|A)$ that the system takes a short step from the macrostate A to the neighboring macrostate $A' = A + dA$. The transition probability $P(A'|A)$ will be assigned by maximizing an entropy. This first requires that we identify the particular entropy that is relevant to our problem. Next, we must decide on the prior distribution: what short steps we might expect before we know the specifics of the motion. Finally, we stipulate the constraints that are meant to capture the information that is relevant to the particular problem at hand.

To settle the first item—the choice of entropy—we note that not only are we are uncertain about the macrostate at A, but we are also uncertain about the microstates $x \in \mathcal{X}$. This means that the actual universe of discourse is the joint space $\mathcal{A} \times \mathcal{X}$ and the appropriate statistical description of the system is in terms of the joint distribution

$$P(x, A) = P(x|A)P(A) = \rho(x|A)P(A) \, , \tag{20}$$

Where ρ is of form (3), which means that we impose $P(x|A)$ to be canonical and the distribution $P(A)$ represents our lack of knowledge about the macrostates. Note that what we did in (20) is nothing more than assuming a probability distribution for the macrostates. This description is sometimes referred to as superstatistics [60].

Our immediate task is to find the transition probability of a change $P(x', A'|x, A)$ by maximizing the entropy

$$\mathsf{S}[P|Q] = -\int dA' dx' P(x', A'|x, A) \log \frac{P(x', A'|x, A)}{Q(x', A'|x, A)} \, , \tag{21}$$

relative to the prior $Q(x', A'|x, A)$ and subject to constraints to be discussed below (to simplify the notation in multidimensional integrals we write $d^n A' = dA'$ and $d^n x' = dx'$).

Although S in (6) and S in (21) are both entropies, in the information theory sense, they represent two very distinct statistical objects. The $S(A)$ in (6) is the entropy of the macrostate—which is what one may be used to from statistical mechanics —while the $\mathsf{S}[P|Q]$ in (21) is the entropy to be maximized, so that we find the transition probability that better matches the information at hand, which means that S is a tool to select the dynamics of the macrostates.

3.2. The Prior

We adopt a prior that implements the idea that the system evolves by taking short steps $A \to A + \Delta A$ at the macrostate level, but is otherwise maximally uninformative. We write

$$Q(x', A'|x, A) = Q(x'|x, A, A')Q(A'|x, A) \, , \tag{22}$$

and analyze the two factors in turn. We shall assume that a priori, before we know the relation between the microstates x and the macrostate A, the prior distribution for x' is the same uniform underlying measure $q(x')$ that is introduced in (1),

$$Q(x'|x, A, A') = q(x') \, . \tag{23}$$

Next, we tackle the second factor $Q(A'|x, A)$. As shown in Appendix A, using the method of maximum entropy, the prior that enforces short steps, but is otherwise maximally uninformative, is spherically symmetric as

$$Q(A'|x, A) = Q(A'|A) \propto g^{1/2}(A') \exp\left[-\frac{1}{2\tau}g_{ij}\Delta A^i \Delta A^j\right]. \tag{24}$$

so the joint prior is

$$Q(x', A'|x, A) \propto q(x')g^{1/2}(A') \exp\left[-\frac{1}{2\tau}g_{ij}\Delta A^i \Delta A^j\right]. \tag{25}$$

We see that steps of length

$$\Delta\ell \sim (g_{ij}\Delta A^i \Delta A^j)^{1/2} \gg \tau^{1/2}, \tag{26}$$

have negligible probability. Eventually, we will take the limit $\tau \to 0$ to enforce short steps. The prefactor $g^{1/2}(A')$ ensures that $Q(A'|A)$ is a probability density. Later, we will show how this choice of priors, which only comes from the assumption of continuous motion, leads to a diffusion structure.

3.3. The Constraints

The piece of information that we wish to codify through the constraints is the simple geometric idea that the dynamics remains confined to the statistical manifold $\mathcal{A}$. This is implemented by writing

$$P(x', A'|x, A) = P(x'|x, A, A')P(A'|x, A) \tag{27}$$

and imposing that the distribution for x' is a canonical distribution

$$P(x'|x, A, A') = \rho(x'|A') \in \mathcal{A}. \tag{28}$$

This means that, given A', the distribution of x' is independent of the initial microstate x and macrostate A. The second factor in (27), $P(A'|x, A)$, is the transition probability we seek, which leads to the constraint

$$P(x', A'|x, A) = \rho(x'|A')P(A'|x, A). \tag{29}$$

We note that this constraint is not, as is usual in applications of the method of maximum entropy, in the form of an expected value. It may appear from (29) that the transition probability $P(A'|x, A)$ will be largely unaffected by the underlying space of microstates. To the contrary, as we shall see below—(31) and (32)—the macrostate dynamics turns out to be dominated by the entropy of the microstate distribution $\rho(x'|A')$.

Depending on the particular system under consideration, one could formulate richer forms of dynamics by imposing additional constraints. To give just one example, one could introduce some drift relative to the direction that is specified by a covector F_i by imposing a constraint of the form $\langle \Delta A^i \rangle F_i = \kappa$ (see [29,30]). However, in this paper, we shall limit ourselves to what is perhaps the simplest case, the minimal ED that is described by the single constraint (29).

3.4. Maximizing the Entropy

Substituting (25) and (29) into (21) and rearranging, we find

$$\mathsf{S}[P|Q] = \int dA' \, P(A'|x, A)\left[-\log\frac{P(A'|x, A)}{Q(A'|A)} + S(A')\right] \tag{30}$$

where $S(A')$ is the macrostate entropy that is given in (6). Maximizing S subject to normalization gives

$$P(A'|x, A) \propto Q(A'|A)e^{S(A')}$$

$$\propto g^{1/2}(A') \exp\left[-\frac{1}{2\tau}g_{ij}\Delta A^i \Delta A^j + S(A')\right]. \tag{31}$$

It is noteworthy that $P(A'|x, A)$ turned out to be independent of x, which is not surprising, since neither the prior nor the constraints indicate any correlation between A' and x.

We perform a linear approximation of S because the transition from A to A' has to be an arbitrarily small continuous change. This makes the exponential factor in (31) quadratic in ΔA, as

$$P(A'|A) = \frac{g^{1/2}(A')}{\mathcal{Z}} \exp\left[\frac{\partial S}{\partial A^i}\Delta A^i - \frac{1}{2\tau}g_{ij}\Delta A^i \Delta A^j\right], \tag{32}$$

where $e^{S(A)}$ was absorbed in the normalization factor $\mathcal{Z}$. This is the transition probability found by maximizing the entropy (21). However, some mathematical difficulties arise from the fact that (32) is defined over a curved manifold. We are going to explore these mathematical issues and their consequences to motion in the following section.

4. The Transition Probability

Because the statistical manifold is a curved space, we must understand how the transition probability (32) behaves under a change of coordinates. Because (25) and (32) describe an arbitrarily small step, we wish to express the transition probability, as well as the quantities derived from it, which are calculated up to the order of τ. Because the exponent in (32) is manifestly invariant, one can complete squares and obtain

$$P(A'|A) = \frac{g^{1/2}(A')}{\mathcal{Z}'} \exp\left[-\frac{1}{2\tau}g_{ij}\left(\Delta A^i - \tau g^{ik}\frac{\partial S}{\partial A^k}\right)\left(\Delta A^j - \tau g^{ik}\frac{\partial S}{\partial A^k}\right)\right]. \tag{33}$$

If $g(A)$ were uniform, it would imply that the first two moments $\langle \Delta A^i \rangle$ and $\langle \Delta A^i \Delta A^j \rangle$ are of order τ. Therefore, even in the limit $\tau \to 0$, the transition will be affected by curvature effects. This can be verified for an arbitrary metric tensor by a direct calculation of the first moment,

$$\langle \Delta A^i \rangle = \int dA' \, \Delta A^i P(A'|A)$$
$$= \frac{1}{\mathcal{Z}'} \int dA' \, g^{1/2}(A')\Delta A^i \exp\left[-\frac{g_{kl}}{2\tau}\left(\Delta A^k - \tau V^k\right)\left(\Delta A^l - \tau V^l\right)\right], \tag{34}$$

where $V^i = g^{ij}\frac{\partial S}{\partial A^j}$. And the second moment

$$\langle \Delta A^i \Delta A^j \rangle = \int dA' \, \Delta A^i \Delta A^j P(A'|A)$$
$$= \frac{1}{\mathcal{Z}'} \int dA' \, g^{1/2}(A')\Delta A^i \Delta A^j \exp\left[-\frac{g_{kl}}{2\tau}\left(\Delta A^k - \tau V^k\right)\left(\Delta A^l - \tau V^l\right)\right]. \tag{35}$$

It is convenient to write (32) in normal coordinates at A in order to facilitate the calculation of the integrals in (34) and (35). This means that, for a smooth manifold, one can always make a change of coordinates $A^\mu(A^i)$—we will label the normal coordinates with Greek letter indexes (μ, ν)—so that the metric in this coordinate system is so that

$$g_{\mu\nu}(A) = \delta_{\mu\nu} \text{ and } \left.\frac{\partial g_{\mu\nu}}{\partial A^\mu}\right|_A = 0, \tag{36}$$

allowing for us to approximate $g(A') = 1$ for a short step. For a general discussion and rigorous proof of the existence of normal coordinates, see [61]. Although normal

coordinates are a valuable tool for geometrical analysis at this point, it is not clear whether they can be given a deeper statistical interpretation—this is unlike other applications of differential geometry, such as general relativity, where the physical interpretation of normal coordinates turns out be of central importance. A displacement in these coordinates ΔA^μ can be related to the original coordinates by a Taylor expansion in terms of ΔA^i as (see [62,63])

$$\Delta A^\mu = \frac{\partial A^\mu}{\partial A^i}\Delta A^i + \frac{1}{2}\frac{\partial^2 A^\mu}{\partial A^j \partial A^k}\Delta A^j \Delta A^k + o(\tau) \,. \tag{37}$$

To proceed, it is interesting to recall the Christoffel symbols Γ^i_{jk},

$$\Gamma^i_{jk} = \frac{1}{2}g^{il}\left(\frac{\partial g_{jl}}{\partial A^l} + \frac{\partial g_{lj}}{\partial A^k} - \frac{\partial g_{jk}}{\partial A^l}\right) \,, \tag{38}$$

which transform as

$$\Gamma^i_{jk} = \frac{\partial A^i}{\partial A^\mu}\frac{\partial A^\nu}{\partial A^j}\frac{\partial A^\sigma}{\partial A^k}\Gamma^\mu_{\nu\sigma} - \frac{\partial A^i}{\partial A^\mu}\frac{\partial^2 A^\mu}{\partial A^j \partial A^k} \,. \tag{39}$$

Because, in normal coordinates, we have $\Gamma^\mu_{\nu\sigma} = 0$, this allows us to isolate ΔA^i up to the order τ obtaining

$$\Delta A^i = \frac{\partial A^i}{\partial A^\mu}\Delta A^\mu - \frac{1}{2}\Gamma^i_{jk}\Delta A^j \Delta A^k \,, \tag{40}$$

By squaring (40), we have

$$\Delta A^i \Delta A^j = \frac{\partial A^i}{\partial A^\mu}\frac{\partial A^j}{\partial A^\nu}\Delta A^\mu \Delta A^\nu + o(\tau) \,. \tag{41}$$

Because the exponent in (34) is invariant and in a coordinate transformation we have $dA\,P(A) = d\tilde{A}\,P(\tilde{A})$, it separates into two terms.

$$
\begin{aligned}
\langle \Delta A^i \rangle &= \frac{\partial A^i}{\partial A^\mu}\frac{1}{Z'}\int dA'\Delta A^\mu \exp\left[-\frac{\delta_{\nu\sigma}}{2\tau}(\Delta A^\nu - \tau V^\nu)(\Delta A^\sigma - \tau V^\sigma)\right] \\
&\quad -\frac{1}{2}\Gamma^i_{jk}\frac{\partial A^j}{\partial A^\mu}\frac{\partial A^k}{\partial A^\nu}\frac{1}{Z'}\int dA'\Delta A^\mu \Delta A^\nu \exp\left[-\frac{\delta_{\nu\sigma}}{2\tau}(\Delta A^\nu - \tau V^\nu)(\Delta A^\sigma - \tau V^\sigma)\right] \,.
\end{aligned} \tag{42}
$$

The integrals can be evaluated from the known properties of a Gaussian. The integral in the first term gives $\langle \Delta A^\mu \rangle = \tau \delta^{\mu\nu}\frac{\partial S}{\partial A^\nu}$ and the integral in the second term gives $\langle \Delta A^\mu \Delta A^\nu \rangle = \tau \delta^{\mu\nu}$, so that

$$\langle \Delta A^i \rangle = \frac{\partial A^i}{\partial A^\mu}\tau\delta^{\mu\nu}\frac{\partial S}{\partial A^\mu} - \frac{1}{2}\Gamma^i_{jk}\frac{\partial A^j}{\partial A^\mu}\frac{\partial A^k}{\partial A^\nu}\tau\delta^{\mu\nu} \,. \tag{43}$$

Therefore, in natural coordinates, the first two moments up to order of τ are

$$\langle \Delta A^i \rangle = \tau g^{ij}\frac{\partial S}{\partial A^j} - \frac{\tau}{2}\Gamma^i \,, \text{ and } \langle \Delta A^i \Delta A^j \rangle = \tau g^{ij} \,, \tag{44}$$

where $\Gamma^i = \Gamma^i_{jk}g^{jk}$. Here, we see the dependence on curvature for $\langle \Delta A^i \rangle$ in the Christoffel symbol term. Note that it is a consequence of the dependance between ΔA^i and the quadratic term $\Delta A^i \Delta A^j$ in (40), which per (44) does not vanish, even for small steps. Hence, fluctuations in A^i cannot be ignored in the ED motion, and this is the reason why the motion probes curvature. It also follows from (44) that, even in the limit $\tau \to 0$, the average change ΔA^i does not transform covariantly.

Note that we used several words, such as "transitions", "short step", "continuous", and "dynamics" without any established notion of time. In the following section, we will discuss time not as an external parameter, but as an emergent parameter from the maximum entropy transition (32) and its moments (44).

5. Entropic Time

Having described a short step transition, the next challenge is to study how these short steps accumulate.

5.1. Introducing Time

In order to introduce time, we note that A' and A are elements of the same manifold; therefore, $P(A')$ and $P(A)$ are two probability distributions over the same space. Our established solution for describing the accumulation of changes (see [28]) is to introduce a "book-keeping" parameter t that distinguishes the said distributions as labelled by different parameters, i.e., $P(A') = P_{t'}(A)$ and $P(A) = P_t(A)$.

In this formalism, we will refer to these different labels as a description of the system at particular instants t and t'. This allows us to call $P(A'|A)$ a transition probability.

$$P_{t'}(A) = P(A') = \int dA \, P_{\Delta t}(A'|A)P_t(A) \tag{45}$$

where $\Delta t = t' - t$.

As the system changes from A to A' and then to A''. The probability $P(A'')$ will be constructed from $P(A')$, not explicitly dependent on $P(A)$. This means that (45) represents a Markovian process: conditioned on the present $P_{t'}(A)$, the "future" $P_{t''}(A)$ is independent of the "past" $P_t(A)$, where $t'' > t' > t$. It is important to notice that, under this formalism, (45) is not used to show that the process is Markovian in an existing time, but rather the concept of time that was developed here makes the dynamics Markovian by design.

It is also important to notice that the parameter t that is presented here is not necessarily the "physical" time (as it appears in Newton's laws of motion or the Schrödinger equation). Our parameter t, which we call entropic time, is an epistemic well-ordered parameter in which the dynamics are defined.

5.2. The Entropic Arrow of Time

It is important to note that the marginalization process from (20) to (45) could also lead to

$$P(A) = \int dA' \, P(A|A')P(A') \,, \tag{46}$$

where the conditional probabilities are related by Bayes' Theorem,

$$P(A|A') = \frac{P(A)}{P(A')}P(A'|A) \,, \tag{47}$$

showing that a change "forward" will not happen the same way as a change "backwards" unless the system is in some form of stationary state, $P(A) = P(A')$. Another way to present this is that probability theory alone gives no intrinsic distinction of the change "forward" and "backward". The fact that we assigned the change "forward" by ME implies that, in general, the change "backward" is not an entropy maximum. Therefore, the preferential direction of the flow of time arises from the entropic dynamics naturally.

5.3. Calibrating the Clock

One needs to define the duration Δt with respect to the motion in order to connect the entropic time to the transition probability. Time in entropic dynamics is defined so as to simplify the description of the motion. This notion of time is tailored to the system under discussion. The time interval will be chosen, so that the parameter τ that first appeared in the prior (25) takes the role of a time interval,

$$\tau = \eta \Delta t \,, \tag{48}$$

where η is a constant, so that t has the units of time. For the remainder of this article, we will adopt $\eta = 1$. In principle, any monotonic function $t(\tau)$ serves as an parameter for

ordering. Our choice is a matter of convenience, as required by simplicity. Here this is implemented so that for a short transition we have the dimensionless time interval

$$\Delta t = g_{ij} \langle \Delta A^i \Delta A^j \rangle \ . \tag{49}$$

This means that the system's fluctuations measure the entropic time. Rather than having the changes in the system represented in terms of given time intervals (as measured by an external clock), here the system is its own clock.

The moments in (44) can be written, up to order Δt, as

$$\frac{\langle \Delta A^i \rangle}{\Delta t} = g^{ij} \frac{\partial S}{\partial A^j} - \frac{1}{2} \Gamma^i \ , \quad \text{and} \quad \frac{\langle \Delta A^i \Delta A^j \rangle}{\Delta t} = g^{ij} \ . \tag{50}$$

With this, we have established a concept of time and it is convenient to write the trajectory of the expected values in terms of a differential equation.

6. Diffusion and the Fokker–Planck Equation

Our goal of designing the dynamics from entropic methods is accomplished. The entropic dynamics equation of evolution is written in integral form as a Chapman–Kolmogorov equation (45) with a transition probability given by (32). In this section, we will conveniently rewrite it in the differential form. The computed drift $\langle \Delta A^i \rangle$ and the fluctuation $\langle \Delta A^i \Delta A^j \rangle$ in (50) describe the dynamical process as a smooth diffusion—meaning, as defined by [63], a stochastic process in which the first two moments are calculated to the order of Δt, $\langle \Delta A^i \rangle = b^i \Delta t$, $\langle \Delta A^i \Delta A^j \rangle = \eta g^{ij} \Delta t$, and $\langle \Delta A^i \Delta A^j \Delta A^k \rangle = 0$. Therefore, for a short transition, it is possible to write the evolution of $P_t(A)$, as a Fokker–Planck (diffusion) equation,

$$\frac{\partial}{\partial t} P = -\frac{\partial}{\partial A^i} \left(P v^i \right) \ , \tag{51}$$

where

$$v^i = g^{ij} \frac{\partial S}{\partial A^j} - \frac{1}{2} g^{ij} \frac{\partial}{\partial A^j} \left(\log \frac{P}{g^{1/2}} \right) \ . \tag{52}$$

The derivation of (51) and (52) takes into account the fact that the space in which the diffusion happens is curved and it is given in Appendix B. In equation (52), we see that the current velocity v^i consists of two components. The first term is the drift velocity that is guided by the entropy gradient and the second term is an osmotic velocity, which is a term that is driven by differences in probability density. The examples that are presented in the following section will show how these terms interact and the dynamical properties that are derived from each.

Derivatives and Divergence

Because the entropy S is a scalar, the velocity that is defined in (52) is a contravariant vector. However, (51) is not a manifestly invariant equation. To check its consistency, it is convenient to write it in terms of the invariant object p, being defined as

$$p(A) = \frac{P(A)}{g^{1/2}(A)} \ , \tag{53}$$

meaning that p is the probability of A divided by the volume element, in terms of which (51) becomes

$$\frac{\partial}{\partial t} p = -\frac{1}{g^{1/2}} \frac{\partial}{\partial A^i} \left(g^{1/2} \, p v^i \right) \ . \tag{54}$$

We can recognize, on the right-hand side, the covariant divergence of the contravariant vector pv^i, which can be written in the manifestly covariant form

$$\frac{\partial}{\partial t} p = -\mathcal{D}_i(pv^i) \, , \tag{55}$$

where $\mathcal{D}_i$ is the covariant derivative. The fact that the covariant derivative arises from the dynamical process is the direct indication that even when evolving the invariant object p the curvature of the space is taken into account. We can identify (55) as a continuity equation—generalized to the parallel transport in a curved space, as evidenced by the covariant divergence—where the flux, $j^i = pv^i$, can be written from (52) and (53) as

$$j^i = pg^{ij}\frac{\partial S}{\partial A^j} - \frac{1}{2}g^{ij}\frac{\partial p}{\partial A^j} \, . \tag{56}$$

The second term, which is related to the osmotic velocity, is a Fick's law with diffusion tensor $D^{ij} = g^{ij}/2$. Note that this is identified from purely probabilistic arguments, rather than assuming a repulsive interaction from the microstate dynamics.

Having the dynamics fully described, we can now study its consequences, as will be done in the following section.

7. Examples

We established the entropic dynamics by finding the transition probability (32), presenting it as a differential equation in (51), (52), and presenting it as the invariant equation (55). We want to show some examples of how it would be applied and what are the results achieved. Our present goal is not to search for realistic models, but to search for models that are both mathematically simple and general enough so it can give insight on how to use the formalism.

We will be particularly interested in two properties: the drift velocity,

$$v^i_D = g^{ij}\frac{\partial S}{\partial A^j} \, , \tag{57}$$

which is the first term in (52), and the static states, $v^i = 0$, which are a particular subset of the dynamical system's equilibrium $\partial_t P = 0$. These are obtained from (52) as

$$v^i = 0 \Rightarrow \frac{\partial S}{\partial A^i} - \frac{1}{2}\frac{\partial}{\partial A^i}\log\left(\frac{P}{g^{1/2}}\right) = 0 \tag{58}$$

allowing for one to write the static probability

$$P(A) \propto g^{1/2}(A)\exp[2S(A)] \, , \tag{59}$$

where the factor of 2 in the exponent comes from the diffusion tensor $D^{ij} = g^{ij}/2$ that is explained in Section 6. This result shows that the invariant stationary probability density (53) is

$$p(A) \propto \exp[2S(A)]. \tag{60}$$

7.1. A Gaussian Manifold

The statistical manifold defined by the mean values and correlations of a random variable, the microstate x, is the space of Gaussian distribution, which is an example of a canonical distribution. Here, we consider the dynamics of a two-dimensional spherically symmetric Gaussian with a non-uniform variance, $\sigma(A) = \sigma(A^1, A^2)$, as defined by

$$\left\langle x^1\right\rangle = A^1, \quad \left\langle x^2\right\rangle = A^2, \quad \text{and} \quad \left\langle (x^i - A^i)(x^j - A^j)\right\rangle = \sigma^2(A)\delta^{ij}. \tag{61}$$

These Gaussians are of the form,

$$\rho(x \mid A) = \frac{1}{2\pi\sigma^2(A)} \exp\left(-\frac{1}{2\sigma^2(A)} \sum_{i=1}^{2} (x^i - A^i)^2\right) \tag{62}$$

The entropy of (62) relative to a uniform background measure is given by

$$S(A) = \log\left(2\pi\sigma(A)^2\right) \tag{63}$$

The space of Gaussians with a uniform variance, $\sigma(A) = $ constant, is flat and the dynamics turn out to be a rather trivial spherically symmetric diffusion. Choosing the variance to be non-uniform yields richer and more interesting dynamics. Because this example is pursued for purely illustrative purposes, we restrict to two dimensions and spherically symmetric Gaussians. The generalization is immediate.

The FRIM for a Gaussian distribution is found using (12) (see also [13]), to be

$$dl^2 = \frac{4}{\sigma^2}(d\sigma)^2 + \frac{\delta_{ij}}{\sigma^2}dA^i dA^j\,, \tag{64}$$

so that, using

$$d\sigma = \frac{\partial\sigma}{\partial A^i}dA^i\,, \tag{65}$$

the induced metric $dl^2 = g_{ij}dA^i dA^j$ leads to

$$g_{ij} = \frac{1}{\sigma^2}\left(4\frac{\partial\sigma}{\partial A^i}\frac{\partial\sigma}{\partial A^j} + \delta_{ij}\right). \tag{66}$$

Gaussian Submanifold around an Entropy Maximum

We present an example of our dynamical model that illustrates the motion around an entropy maximum. A simple way to manifest it is to recognize that, in (52), $-S$ plays a role analogous to a potential. A rotationally symmetric quadratic potential can then be sustituted in (63), leading to

$$\sigma(A) = \exp\left(-\frac{(A^1)^2 + (A^2)^2}{4}\right), \tag{67}$$

which, substituted in (66), yields the metric

$$g_{ij} = \begin{bmatrix} (A^1)^2 + \sigma^{-2} & A^1 A^2 \\ A^1 A^2 & (A^2)^2 + \sigma^{-2} \end{bmatrix}, \tag{68}$$

so that

$$g^{1/2} = \sqrt{\left[(A^1)^2 + (A^2)^2\right]\sigma^{-2} + \sigma^{-4}}\,. \tag{69}$$

The scalar curvature for the Gaussian submanifold can be calculated from (68) as

$$R = \frac{\phi^2 - 2\phi}{(\phi^2 + \sigma^{-2})^2}\sigma^2, \quad \text{where} \quad \phi = (A^1)^2 + (A^2)^2\,. \tag{70}$$

From (57), the drift velocity (Figure 1) is

$$v_D^1 = -\frac{A^1\sigma^{-2}}{g} \quad \text{and} \quad v_D^2 = -\frac{A^2\sigma^{-2}}{g}\,. \tag{71}$$

and, from (59), the static probability (Figure 2) is

$$P(A) \propto 4\pi^2 g^{1/2}\sigma^4\,. \tag{72}$$

The static distribution results from the dynamical equilibrium between two opposite tendencies. One is the drift velocity field that drives the distribution along the entropy gradient towards the entropy maximum at the origin. The other is the osmotic diffusive force that we identified earlier as the ED analogue of Fick's law. This osmotic force drives the distribution against the direction of the probability gradient and prevents it from becoming infinitely concentrated at the origin. At equilibrium, the cancellation between these two opposing forces results in the Gaussian distribution, Equation (72).
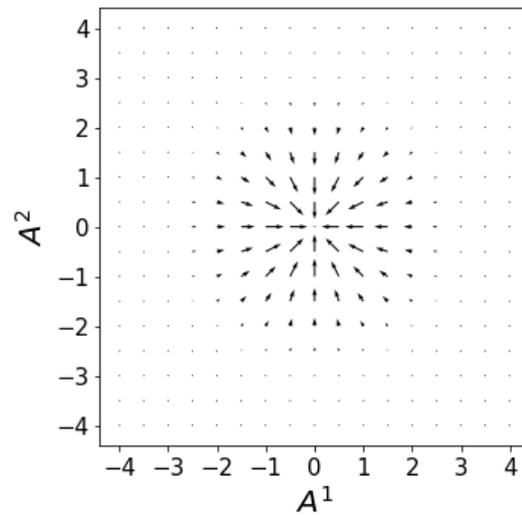
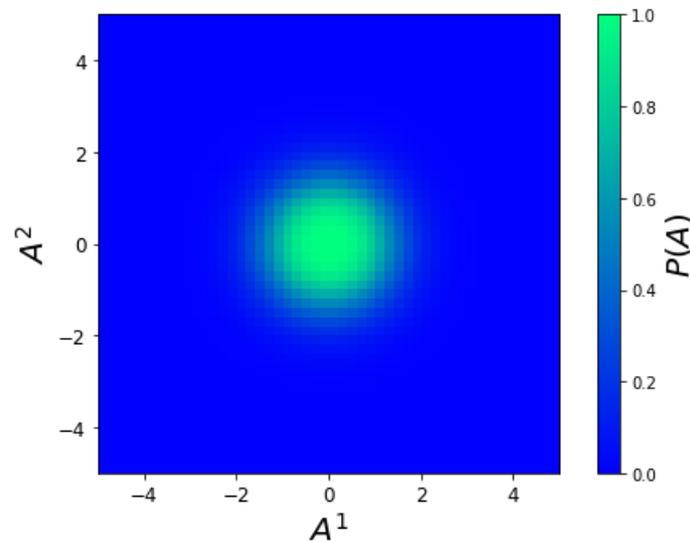**Figure 1.** The drift velocity field (71) drives the flux along the entropy gradient.

Figure 2. Equilibrium stationary probability (72).

7.2. 2-Simplex Manifold

Here, we discuss an example of discrete microstates. The macrostate coordinates, being expected values, are continuous variables. Our subject matter will be a three-state system, $x \in \{1, 2, 3\}$, such as, for example, a 3-sided die. The statistical manifold is the 2-dimensional simplex and the natural coordinates are the probabilities themselves,

$$\mathcal{S}_2 = \left\{ \rho(x) \mid \rho(x) \geq 0 \,, \ \sum_{x=1}^{3} \rho(x) = 1 \right\}. \tag{73}$$

The distributions on the two-simplex are Gibbs distributions defined by the sufficient statistics of functions

$$a^i(x) = \delta^i_x \quad \text{so that} \quad A^i = \langle a^i \rangle = \rho(i) . \tag{74}$$

The entropy relative to the uniform discrete measure is

$$S = -\sum_{i=1}^{3} \rho(i) \log(\rho(i)) = -\sum_{i=1}^{3} A^i \log\left(A^i\right) , \tag{75}$$

and the information metric is given by

$$g_{ij} = \sum_{k=1}^{3} \rho^k \frac{\partial \log\left(\rho^k\right)}{\partial A^i} \frac{\partial \log\left(\rho^k\right)}{\partial A^j} . \tag{76}$$

The two-simplex arises naturally from probability theory due to normalization when one identifies the macrostate of interest to be the probabilities themselves. The choice of sufficient statistics (74) implies that the manifold is a two-dimensional surface, since, due to the normalization, one can write $A^3 = 1 - A^1 - A^2$. We will use the the tuple (A^1, A^2) as our coordinates and A^3 as a function of them. In this scenario, one finds a metric tensor

$$g_{ij} = \begin{bmatrix} \dfrac{1}{A^3} + \dfrac{1}{A^1} & \dfrac{1}{A^3} \\ \dfrac{1}{A^3} & \dfrac{1}{A^3} + \dfrac{1}{A^2} \end{bmatrix} , \tag{77}$$
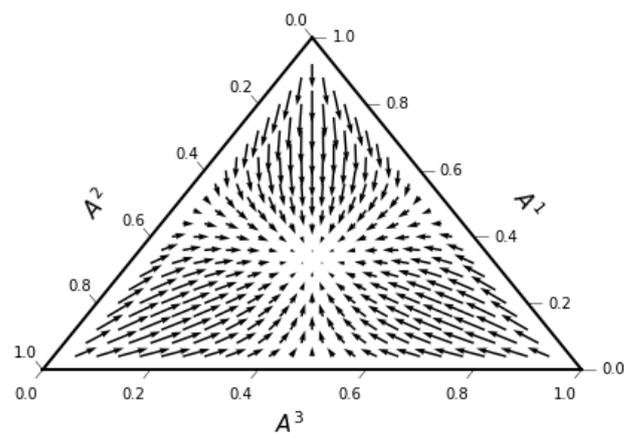
which induces the volume element

$$g^{1/2} = \sqrt{\frac{1}{A^1 A^2 A^3}} . \tag{78}$$

As is well known, the simplex is characterized by a constant curvature $R = 1/2$; the two-simplex is the positive octant of a sphere. From (57), the drift velocity (Figure 3) is

$$\begin{aligned} v_D^1 &= A^1 \left[A^2 \log\left(\frac{A^2}{A^3}\right) + (A^1 - 1) \log\left(\frac{A^1}{A^3}\right) \right] \\ v_D^2 &= A^2 \left[A^1 \log\left(\frac{A^1}{A^3}\right) + (A^2 - 1) \log\left(\frac{A^2}{A^3}\right) \right] , \end{aligned} \tag{79}$$

Additionally, the static probability is

$$P(A) \propto g^{1/2} \prod_{i=1}^{3} \left(A^i\right)^{-2A^i} . \tag{80}$$

Figure 3. Drift velocity field for the two-simplex in (79). The ternary plots ware created using python-ternary library [64].

From the determinant of the metric, we note that the static probability (80) diverges at the boundary of the two-simplex. This reflects the fact that a two-state system (say, $i = 1, 2$) is easily distinguishable from a three-state system ($i = 1, 2, 3$). Indeed, a single datum $i = 3$ will tell us that we are dealing with a three-state system.

On the other hand, we can see (Figure 4) that this divergence is not present in the invariant stationary probability (53).
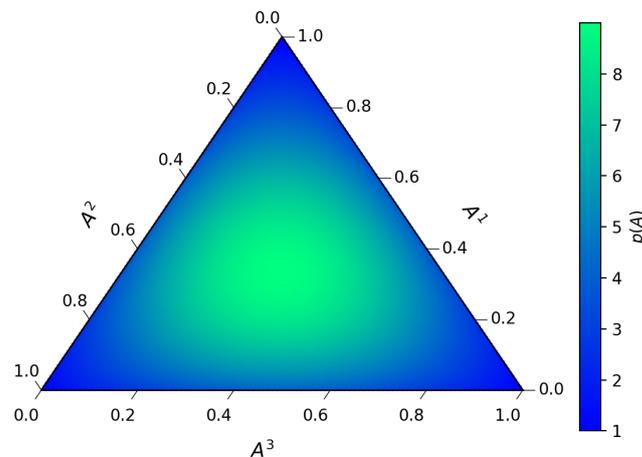
Figure 4. Static invariant stationary probability for the three-state system.

As in the Gaussian case discussed in the previous section, the static equilibrium results from the cancellation of two opposing forces: the entropic force along the drift velocity field towards the center of the simplex is cancelled by the osmotic diffusive force away from the center.

8. Conclusions

We conclude with a summary of the main results. In this paper, the entropic dynamics framework has been extended to describe dynamics on a statistical manifold. ME played an instrumental role in that it allowed us to impose constraints that are not in the standard form of expected values.

The resulting dynamics, which follow from purely entropic considerations, take the form of a diffusive process on a curved space. The effects of curvature turn out to be significant. We found that the probability flux is the result of two components. One describes a flux along the entropy gradient and the other is a diffusive or osmotic

component that turns out to be the curved-space analogue of Fick's law with a diffusion tensor $D^{ij} = g^{ij}/2$ that is given by information geometry.

A highlight of the model is that it includes an "entropic" notion of time that is tailored to the system under study; the system is its own clock. This opens the door to the introduction of a notion of time that transcends physics and it might be useful for social and ecological systems. The emerging notion of entropic time is intrinsically directional. There is a natural arrow of time that manifests itself in a simple description of the approach to equilibrium.

The model developed here is rather minimal in the sense that the dynamics could be extended by taking additional relevant information into account. For example, it is rather straightforward to enrich the dynamics by imposing additional constraints

$$\langle \Delta A^i \rangle F_i(A) = \kappa' \,, \tag{81}$$

involving system-specific functions $F_i(A)$ that carry information regarding correlations. This is the kind of further developments that we envisage in future work.

As illustrative examples, the dynamics were applied to two general spaces of probability distributions. A submanifold of the space of two-dimensional Gaussians and the space of probability distributions for a three-state system (two-simplex). In each of these, we were able to provide insight on the dynamics by presenting the drift velocity (57) and the equilibrium stationary states (59). Additionally, as future work, we intend to apply the dynamics developed here in the distributions found in network sciences [65].

Abbreviations

The following abbreviations are used in this manuscript:

MaxEnt	MAXimum ENTropy—according to Jaynes [2]
ME	(Modern) Maximum Entropy—since Shore & Johnson [7]
FRIM	Fisher-Rao Information Metric
ED	Entropic Dynamics

Appendix A. Obtaining the Prior

In this appendix we derive the prior transition probability from A to A' seen in (25). This is achieved by maximizing the entropy

$$S[Q] = -\int dA' \, Q(A'|x, A) \log \left(\frac{Q(A'|x, A)}{R(A'|x, A)} \right) \,, \tag{A1}$$

where $R(A'|x, A)$, the prior for (A1), encodes information about an unbiased transition of the systems. The posterior of (A1), $Q(A'|x, A)$, becomes the prior in (21).

At this stage A could evolve into any A' and the only assumption is that the assigned prior for (A1) leads to equal probabilities for equal volumes; thus, ignoring biases. That

is achieved by a prior proportional to the volume element $R(A'|x, A) \propto g^{1/2}(A')$, where $g(A) = \det g_{ij}(A)$. There is no need to address the normalization of R since it will have no effect on the posterior.

The chosen constraint represents an isotropic and continuous motion on the manifold. This will be imposed by

$$\int dA' \, Q(A'|x, A) \, g_{ij}\Delta A^i \Delta A^j = K \, . \tag{A2}$$

where K is a small quantity, since $g_{ij}\Delta A^i \Delta A^j$ is invariant only in the limit for short steps $\Delta A^i \to 0$. Therefore, eventually, $K \to 0$.

The result of maximizing (A1) under (A2) and normalization is

$$Q(A'|x, A) \propto g^{1/2}(A') \exp\left(-\alpha \, g_{ij}\Delta A^i \Delta A^j\right) \, , \tag{A3}$$

where α is the Lagrange multiplier associated to (A2). As the result requires $K \to 0$ to make it geometrically invariant, the conjugated Lagrange multiplier should be allowed to be taken to infinity. This allows us to define $\tau = 1/\alpha$, such that the short step limit will lead to $\tau \to 0$.

Note that, since no motion in x and no correlation between x and A' is induced by the constraints, the result does not depend on the previous microstate x, $Q(A'|x, A) = Q(A'|A)$.

Appendix B. Derivation of the Fokker-Planck Equation

The goal of this appendix is to show that for a dynamics that is a smooth diffusion in a curved space, can be written as a Fokker-Planck equation and to obtain its velocity (52) from the moments for the motion (50). In order to do so, it is convenient to define a drift velocity

$$b^i = \lim_{\Delta t \to 0} \frac{\langle \Delta A^i \rangle}{\Delta t} = g^{ij}\frac{\partial S}{\partial A^j} - \frac{1}{2}\Gamma^i \, . \tag{A4}$$

First, let us analyze the change of a smooth integrable function $f(A)$ as the system transitions from A to A'. A smooth change in the function $f(A)$ will be

$$\Delta f(A) = \frac{\partial f}{\partial A^i}\Delta A^i + \frac{1}{2}\frac{\partial^2 f}{\partial A^i \partial A^j}\Delta A^i \Delta A^j + o(\Delta t) \, , \tag{A5}$$

since a cubic term, $\Delta A^i \Delta A^j \Delta A^k$ would be $o(\Delta t)$. In a smooth diffusion we can take the expected value of (A5) with respect to $P(A'|A)$ as

$$\langle \Delta f(A) \rangle = \int dA' \, P(A'|A)(f(A') - f(A)) = \left(b^i\frac{\partial}{\partial A^i} + \frac{1}{2}g^{ij}\frac{\partial^2}{\partial A^i \partial A^j}\right)f(A)\Delta t \, . \tag{A6}$$

which can be further averaged in $P(A)$. The left-hand side will be

$$\int dA \, P(A) \int dA' \, P(A'|A)(f(A') - f(A)) = \int dA dA' P(A', A)(f(A') - f(A)) \tag{A7}$$

$$= \int dA' \, P(A')f(A') - \int dA \, P(A)f(A) \tag{A8}$$

while the right hand is

$$\int dA \, P(A)\left(b^i\frac{\partial}{\partial A^i} + \frac{1}{2}g^{ij}\frac{\partial^2}{\partial A^i \partial A^j}\right)f(A)\Delta t \, . \tag{A9}$$

such that they equate to

$$\int dA' \, P(A')f(A') - \int dA \, P(A)f(A) = \int dA \, P(A)\left(b^i\frac{\partial}{\partial A^i} + \frac{1}{2}g^{ij}\frac{\partial^2}{\partial A^i\partial A^j}\right)f(A)\Delta t \,. \quad \text{(A10)}$$

As established in Section 5, $P(A)$ and $P(A')$ are distributions at the instants t and t' respectively.

$$\int dA \left(\frac{P_{t'}(A) - P_t(A)}{\Delta t}\right)f(A) = \int dA \, P(A)\left(b^i\frac{\partial}{\partial A^i} + \frac{1}{2}g^{ij}\frac{\partial^2}{\partial A^i\partial A^j}\right)f(A)\Delta t \,, \quad \text{(A11)}$$

which can be partially integrated in the limit of small steps

$$\int dA \left(\frac{\partial P(A)}{\partial t}\right)f(A) = \int dA \left(-\frac{\partial}{\partial A^i}(b^i P(A)) + \frac{1}{2}\frac{\partial^2}{\partial A^i\partial A^j}(g^{ij}P(A))\right)f(A) \,. \quad \text{(A12)}$$

Due to the generality of f as test function, we identify the integrants,

$$\frac{\partial}{\partial t}P(A) = -\frac{\partial}{\partial A^i}\left(b^i P(A) - \frac{1}{2}\frac{\partial}{\partial A^j}(g^{ij}P(A))\right) \,, \quad \text{(A13)}$$

and substitute b^i (A4) for general coordinates,

$$\frac{\partial}{\partial t}P(A) = -\frac{\partial}{\partial A^i}\left(g^{ij}\frac{\partial S}{\partial A^j}P(A) - \frac{1}{2}\Gamma^i P(A) - \frac{1}{2}\left(\frac{\partial g^{ij}}{\partial A^j}\right)P(A) - \frac{1}{2}g^{ij}\frac{\partial P(A)}{\partial A^j}\right) \,, \quad \text{(A14)}$$

and the contracted Christoffel symbols can be substituted in the identity

$$\Gamma^i = -\frac{1}{g^{1/2}}\frac{\partial}{\partial A^j}(g^{1/2}g^{ij}) = -\frac{\partial g^{ij}}{\partial A^j} - g^{ij}\frac{\partial \log g^{1/2}}{\partial A^j} \,. \quad \text{(A15)}$$

Here we see that the effect of curvature—encoded by the Christoffel symbols—substitute in the differential Equation (A13) obtaining

$$\frac{\partial}{\partial t}P(A) = -\frac{\partial}{\partial A^i}\left(g^{ij}\frac{\partial S}{\partial A^j} - \frac{1}{2}g^{ij}\frac{\partial}{\partial A^j}\left(\log\frac{P(A)}{g^{1/2}}\right)\right)P(A) \,, \quad \text{(A16)}$$

where the second term inside the parenthesis above is the result of taking the curvature into account. The result is a Fokker-Planck equation that is usefully written in the continuity form

$$\frac{\partial}{\partial t}P = -\frac{\partial}{\partial A^i}\left(Pv^i\right) \,, \quad \text{(A17)}$$

where

$$v^i = g^{ij}\frac{\partial S}{\partial A^j} - \frac{1}{2}g^{ij}\frac{\partial}{\partial A^j}\left(\log\frac{P}{g^{1/2}}\right) \,, \quad \text{(A18)}$$

completing the derivation.

References

1. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
2. Jaynes, E.T. Information theory and statistical mechanics: I. *Phys. Rev.* **1957**, *106*, 620. [CrossRef]
3. Jaynes, E.T. Information theory and statistical mechanics. II. *Phys. Rev.* **1957**, *108*, 171. [CrossRef]
4. Rosenkrantz, R.D. (Ed.) *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*; Reidel: Dordrecht, The Netherlands, 1983. [CrossRef]
5. Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, UK, 2003.
6. Gibbs, J. *Elementary Principles in Statistical Mechanics*; Yale University Press: New Haven, Connecticut, 1902; Reprinted by Ox Bow Press: Woodbridge, Connecticut, 1981.

7.	Shore, J.; Johnson, R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory* **1980**, *26*, 26–37. [CrossRef]

8.	Skilling, J. The Axioms of Maximum Entropy. In *Maximum-Entropy and Bayesian Methods in Science and Engineering*; Erickson, G.J., Smith, C.R., Eds.; Springer: Dordrecht, The Netherlands, 1988; Volumes 31–32, pp. 173–187. [CrossRef]

9.	Caticha, A. Relative Entropy and Inductive Inference. *AIP Conf. Proc. Am. Inst. Phys.* **2004**, *707*, 75–96. [CrossRef]

10.	Caticha, A. Information and Entropy. *AIP Conf. Proc. Am. Inst. Phys.* **2007**, *954*, 11–22. [CrossRef]

11.	Caticha, A.; Giffin, A. Updating Probabilities. *AIP Conf. Proc. Am. Inst. Phys.* **2006**, *872*, 31–42. [CrossRef]

12.	Vanslette, K. Entropic Updating of Probabilities and Density Matrices. *Entropy* **2017**, *19*, 664. [CrossRef]

13.	Caticha, A. *Entropic Physics: Probability, Entropy, and the Foundations of Physics*. Available online: https://www.albany.edu/physics/faculty/ariel-caticha (accessed on 19 April 2021).

14.	Caticha, A.; Golan, A. An entropic framework for modeling economies. *Phys. A Stat. Mech. Appl.* **2014**, *408*, 149–163. [CrossRef]

15.	Harte, J. *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*; OUP Oxford: Oxford, UK, 2011.

16.	Banavar, J.R.; Maritan, A.; Volkov, I. Applications of the principle of maximum entropy: From physics to ecology. *J. Phys. Condens. Matter* **2010**, *22*, 063101. [CrossRef]

17.	De Martino, A.; De Martino, D. An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon* **2018**, *4*, e00596. [CrossRef] [PubMed]

18.	Dixit, P.D.; Lyashenko, E.; Niepel, M.; Vitkup, D. Maximum entropy framework for predictive inference of cell population heterogeneity and responses in signaling networks. *Cell Syst.* **2020**, *10*, 204–212. [CrossRef]

19.	Cimini, G.; Squartini, T.; Saracco, F.; Garlaschelli, D.; Gabrielli, A.; Caldarelli, G. The statistical physics of real-world networks. *Nat. Rev. Phys.* **2019**, *1*, 58–71. [CrossRef]

20.	Radicchi, F.; Krioukov, D.; Hartle, H.; Bianconi, G. Classical information theory of networks. *J. Phys. Complex.* **2020**, *1*, 025001. [CrossRef]

21.	Vicente, R.; Susemihl, A.; Jericó, J.P.; Caticha, N. Moral foundations in an interacting neural networks society: A statistical mechanics analysis. *Phys. A Stat. Mech. Its Appl.* **2014**, *400*, 124–138. [CrossRef]

22.	Alves, F.; Caticha, N. *Sympatric Multiculturalism in Opinion Models*; AIP Conference Proceedings; AIP Publishing LLC.: New York, NY, USA, 2016; Volume 1757, p. 060005. [CrossRef]

23.	Jaynes, E.T. Where do we stand on maximum entropy? In *The Maximum Entropy Principle*; Levine, R.D., Tribus, M., Eds.; MIT Press: Cambridge, MA, USA, 1979. [CrossRef]

24.	Balian, R. *From Microphysics to Macrophysics: Methods and Applications of Statistical Mechanics. Volumes I and II*; Springer: Heidelberg, Germany, 1991–1992.

25.	Pressé, S.; Ghosh, K.; Lee, J.; Dill, K.A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **2013**, *85*, 1115–1141. [CrossRef]

26.	Davis, S.; González, D. Hamiltonian formalism and path entropy maximization. *J. Phys. A Math. Theor.* **2015**, *48*, 425003. [CrossRef]

27.	Cafaro, C.; Ali, S.A. Maximum caliber inference and the stochastic Ising model. *Phys. Rev. E* **2016**, *94*. [CrossRef]

28.	Caticha, A. Entropic dynamics, time and quantum theory. *J. Phys. A Math. Theor.* **2011**, *44*, 225303. [CrossRef]

29.	Caticha, A. The Entropic Dynamics Approach to Quantum Mechanics. *Entropy* **2019**, *21*, 943. [CrossRef]

30.	Ipek, S.; Abedi, M.; Caticha, A. Entropic dynamics: Reconstructing quantum field theory in curved space-time. *Class. Quantum Gravity* **2019**, *36*, 205013. [CrossRef]

31.	Pessoa, P.; Caticha, A. Exact renormalization groups as a form of entropic dynamics. *Entropy* **2018**, *20*, 25. [CrossRef]

32.	Abedi, M.; Bartolomeo, D. Entropic Dynamics of Exchange Rates and Options. *Entropy* **2019**, *21*, 586. [CrossRef] [PubMed]

33.	Abedi, M.; Bartolomeo, D. Entropic Dynamics of Stocks and European Options. *Entropy* **2019**, *21*, 765. [CrossRef]

34.	Caticha, N. Entropic Dynamics in Neural Networks, the Renormalization Group and the Hamilton-Jacobi-Bellman Equation. *Entropy* **2020**, *22*, 587. [CrossRef]

35.	Fisher, R.A. Theory of Statistical Estimation. *Proc. Camb. Philos. Soc.* **1925**, *122*, 700. [CrossRef]

36.	Rao, C.R. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81. [CrossRef]

37.	Amari, S.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Society: Providence, RI, USA, 2000.

38.	Amari, S. *Information Geometry and Its Applications*; Springer International Publishing: Berlin, Germany, 2016. [CrossRef]

39.	Ay, N.; Jost, J.; Lê, H.V.; Schwachhöfer, L. *Information Geometry*; Springer International Publishing: Berlin, Germany, 2017. [CrossRef]

40.	Caticha, A. The basics of information geometry. *AIP Conf. Proc. Am. Inst. Phys.* **2015**, *1641*, 15–26. [CrossRef]

41.	Nielsen, F.; Garcia, V. Statistical exponential families: A digest with flash cards. *arXiv* **2009**, arXiv:cs.LG/0911.4863.

42.	Ruppeiner, G. Riemannian geometry in thermodynamic fluctuation theory. *Rev. Mod. Phys.* **1995**, *67*, 605. [CrossRef]

43.	Janyszek, H.; Mrugala, R. Riemannian geometry and stability of ideal quantum gases. *J. Phys. A Math. Gen.* **1990**, *23*, 467. [CrossRef]

44.	Brody, D.; Rivier, N. Geometrical aspects of statistical mechanics. *Phys. Rev. E* **1995**, *51*, 1006. [CrossRef]

45.	Oshima, H.; Obata, T.; Hara, H. Riemann scalar curvature of ideal quantum gases obeying Gentiles statistics. *J. Phys. A Math. Gen.* **1999**, *32*, 6373–6383. [CrossRef]

46.	Brody, D.; Hook, D.W. Information geometry in vapour–liquid equilibrium. *J. Phys. A Math. Theor.* **2008**, *42*, 023001. [CrossRef]

47.	Yapage, N.; Nagaoka, H. An information geometrical approach to the mean-field approximation for quantum Ising spin models. *J. Phys. A Math. Theor.* **2008**, *41*, 065005. [CrossRef]

48. Tanaka, S. Information geometrical characterization of the Onsager-Machlup process. *Chem. Phys. Lett.* **2017**, *689*, 152–155. [CrossRef]
49. Nicholson, S.B.; del Campo, A.; Green, J.R. Nonequilibrium uncertainty principle from information geometry. *Phys. Rev. E* **2018**, *98*, 032106. [CrossRef]
50. Ay, N.; Olbrich, E.; Bertschinger, N.; Jost, J. A geometric approach to complexity. *Chaos Interdiscip. J. Nonlinear Sci.* **2011**, *21*, 037103. [CrossRef] [PubMed]
51. Felice, D.; Mancini, S.; Pettini, M. Quantifying networks complexity from information geometry viewpoint. *J. Math. Phys.* **2014**, *55*, 043505. [CrossRef]
52. Felice, D.; Cafaro, C.; Mancini, S. Information geometric methods for complexity. *Chaos Interdiscip. J. Nonlinear Sci.* **2018**, *28*, 032101. [CrossRef]
53. Fisher, R.A. On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc. Lond.* **1922**, *222*, 309–368. [CrossRef]
54. Pitman, E.J.G. Sufficient statistics and intrinsic accuracy. *Mathematical Proceedings of the Cambridge Philosophical Society*; Cambridge University Press: Cambridge, UK, 1936; Volume 32, pp. 567–579. [CrossRef]
55. Darmois, G. Sur les lois de probabilitéa estimation exhaustive. *CR Acad. Sci. Paris* **1935**, *260*, 85.
56. Koopman, B.O. On distributions admitting a sufficient statistic. *Trans. Am. Math. Soc.* **1936**, *39*, 399–409. [CrossRef]
57. Brody, D. A note on exponential families of distributions. *J. Phys. A Math. Theor.* **2007**, *40*, F691. [CrossRef]
58. Cencov, N.N. Statistical decision rules and optimal inference. *Am. Math. Soc.* **1981**, *53*. [CrossRef]
59. Campbell, L.L. An extended Cencov characterization of the information metric. *Proc. Am. Math. Soc.* **1986**, *98*, 135–141. [CrossRef]
60. Beck, C.; Cohen, E.G.D. Superstatistics. *Phys. A Stat. Mech. Appl.* **2003**, *322*, 267–275. [CrossRef]
61. Kobayashi, S.; Nomizu, K. *Foundations of Differential Geometry (Wiley Classics Library)*; John Wiley and Sons: New York, NY, USA, 1963; Volume 1.
62. Nawaz, S.; Abedi, M.; Caticha, A. *Entropic Dynamics on Curved Spaces*; AIP Conference Proceedings; AIP Publishing LLC.: New York, NY, USA, 2016; Volume 1757, p. 030004. [CrossRef]
63. Nelson, E. *Quantum Fluctuations*; Princeton University Press: Princeton, NJ, USA, 1985.
64. Python-ternary: Ternary Plots in Python. GitHub Repository. Available online: https://github.com/marcharper/python-ternary/ (accessed on 19 April 2021).
65. Costa, F.X.; Pessoa, P. Entropic dynamics of networks. *Northeast J. Complex Syst.* **2021**, *3*, 5. [CrossRef]