INFORMATION GEOMETRY AND EVOLUTIONARY GAME THEORY

MARC HARPER

ABSTRACT. The Shahshahani geometry of evolutionary game theory is realized as the information geometry of the simplex, deriving from the Fisher information metric of the manifold of categorical probability distributions. Some essential concepts in evolutionary game theory are realized information-theoretically. Results are extended to the Lotka-Volterra equation and to multiple population systems.

1. INTRODUCTION

The replicator equation is a widely-used model of natural selection. This paper explains the realization of the geometry of evolutionary game theory in terms of information theoretical principles, giving a purely mathematical and statistical origination of the replicator equation. Under this interpretation, the replicator equation models the information dynamics of a population of replicating entities. Additionally, the Kullback-Liebler information divergence, which serves as a Lyapunov function for the replicator dynamic, can be interpreted as a measure potential information, characterizing the concept of evolutionary stable state informatically.

1.1. **Continuous Replicator Dynamic.** Consider a categorical distribution X on n categories of entities in a population. This is discrete probability distribution represented by a unit vector of n variables $x = (x_1, \ldots, x_n)$ under the normalization $|x| = x_1 + \cdots + x_n = 1$, where x_i denotes the proportion of the i-th type in the population. The replicator equation on this distribution is the differential equation

$$\dot{x}_i = x_i \left(f_i(x) - \bar{f}(x) \right),$$

where $f(x) = (f_1(x), \ldots, f_n(x))$ is a fitness landscape and $\bar{f}(x) = x_1 f_1(x) + \cdots + x_n f_n(x)$ is the mean fitness.

2. Geometric Aspects of Evolutionary Game Theory

The information theoretic interpretations of the previous chapter have a unified basis in information geometry. We begin with a description of the geometry of the simplex and geometric results known in evolutionary game theory.

2.1. **The Geometry of the Simplex.** Let S^n be the interior of the n-simplex Δ^n, which is $(n-1)$-dimensional. Each point x of the simplex has the property that $x_1 + \cdots + x_n = 1$, so the tangent space at any point on the interior is the $(n-1)$-dimensional vector space described by n vectors $v_1, \ldots, v_n$ such that $v_1 + \cdots + v_n = 0$. The orthogonal complement of the tangent space is the one dimensional line with direction vector $\mathbf{1} = (1, 1, \ldots, 1)$. Indeed $\mathbf{1} \cdot v = 0$ for any v in the tangent space and the complement is necessarily one-dimensional. The faces of Δ^n are isomorphic to a simplex of one lower dimension, which can be seen by setting one of the x_i to zero and indicates the absence of that type in the population. The replicator equation is *forward-invariant* on the simplex (and hence each of its faces), since if $x_i = 0$ then $\dot{x}_i = 0$. Because of this property, the replicator equation is called *non-innovative* since new types cannot arise, in contrast to evolutionary dynamics in which this is possible (notably the replicator-mutator equation [10] and the orthogonal projection dynamic [12]).

2.2. **Shahshahani Geometry.** Shahshahani introduced two Riemannian manifolds into mathematical biology [13]: the positive orthant of $\mathbb{R}^n$, denoted $\mathbb{R}^n_+$, with the metric

$$g_{ij}(x) = \frac{|x|}{x_i} \delta_{ij},$$

where $|x| = \sum_i x_i$ and the restriction to the simplex $\Delta^n = \{x \in \mathbb{R}^n \mid |x| = 1, x_i \geq 0 \,\forall i\} \subset \mathbb{R}^n_+$, with the metric

$$g_{ij}(x) = \frac{1}{x_i} \delta_{ij}.$$

Call the latter manifold the Shahshahani manifold; its metric is known as the Shahshahani metric. There is a normalization map $N : \mathbb{R}^n_+ \to \Delta^n$ given by $x \mapsto \frac{x}{|x|}$. For each $\tau \in \mathbb{R}_+$, there is a map φ_τ mapping the simplex into $\mathbb{R}^n_+$ by $x \mapsto \tau x$. These maps are sections of the normalization map since $N \circ \varphi_\tau = \mathrm{id}_{\Delta^n}$. The Shahshahani metric diverges on the boundary of the simplex so the metric is valid only on the interior S^n.

Dynamics that are forward-invariant, such as the replicator dynamic, are not affected by the discontinuity at the boundary.

2.3. **The Replicator Dynamic, Geometrically.** The geometry of the Shahshahani manifold yields an elegant interpretation of the replicator equation: it is the gradient flow of the Shahshahani metric. Shahshahani proved the result for a special case of the replicator equation; the following more general theorem comes from [8].

Theorem 1. *If the differential equation $\dot{x}_i = f_i(x)$ is a Euclidean gradient with $f_i = \frac{\partial V}{\partial x_i}$ then the replicator equation $x_i = \hat{f}_i(x) = x_i(f_i(x) - \bar{f}(x))$ is a gradient with respect to the Shahshahani metric.*

In the case that the fitness landscape is a Euclidean gradient, the Shahshahani gradient gives a Lyapunov function for the dynamic. The classical case is that of a symmetric matrix A and fitness landscape $f(x) = Ax$, where A is the matrix of Malthusian fitness parameters given by the difference in birth rates and death rates $a_{ij} = b_{ij} - d_{ij}$ of an individual having alleles i and j, where the alleles are of a single gene locus. In this case the Shahshahani potential is the mean fitness $\frac{1}{2}x \cdot f(x) = \frac{1}{2}x \cdot Ax$, with Ax the Euclidean gradient [8, 13].

2.4. **Fisher's Fundamental Theorem and Kimura's Maximal Principle.** Fisher's fundamental theorem is a consequence of the geometric approach.

Theorem 2. *The rate of change of the Shahshahani potential is equal to the variance of the fitness landscape [8].*

$$\dot{V}(x) = Var_x[f(x)].$$

This is a general version of Fisher's Fundamental Theorem of Natural Selection, specializing to the traditional result in the case of a symmetric and linear fitness landscape [8]. Kimura's maximal principle follows from the fact that the replicator equation is a gradient flow [13]. As these are both important results in mathematical biology emerging from the geometry, an interpretation is desired of the Shahshahani metric that provides intuition for the introduced geometry on the simplex in the context of modeling natural selection.

3. The Information Geometry of Natural Selection

An intuitive interpretation of the Shahshahani geometry is provided by information theory. Information geometry [5] studies manifolds of probability distributions $p(s, x)$ on a set S depending on parameters x,

which are the coordinates of the manifold. The manifold is endowed with the Fisher information metric,

$$g_{ij}(x) = \mathbb{E}\left[\frac{\partial \log p}{\partial x^i}\frac{\partial \log p}{\partial x^j}\right]$$

which can be shown to be the unique (up to a constant) metric respecting sufficient statistics [7].

3.1. **The Fisher Information Metric is the Shahshahani Metric.** The manifold of immediate interest is $P(X)$, the set of categorical probability distributions on a finite set X, with the Fisher information metric. In this case, it is convenient to abuse notation by allowing the parameters x and distribution variables s to have the same symbolic representation. [1] There is a natural mapping $\varphi : P(X) \to \Delta^n$, where $|X| = n$, given by $p \to (p(1), p(2), \ldots, p(n)) = (x_1, \ldots, x_n)$. This maps $P(X)$ isometrically onto the simplex, and an easy computation shows that that the Fisher information metric is induced by the Shahshahani metric under this mapping. Simply observe that

$$g_{ij}(x) = \mathbb{E}\left[\frac{\partial \log x}{\partial x^i}\frac{\partial \log x}{\partial x^j}\right] = \sum_k x_k \frac{1}{x_i}\delta_{ik}\frac{1}{x_j}\delta_{jk} = \frac{1}{x_i}\delta_{ij}$$

This result was recognized in [4] and [6].

3.2. **Fisher's Fundamental Theorem.** Fisher's fundamental theorem is built into the geometry of $P(X)$ [5]. Define the maps $E[g]$ on $P(X)$ by $p \mapsto E_p[g]$, where g is from the set of functions $\mathbb{R}^X = \{g : X \to \mathbb{R}\}$ and E_p is the mean taken at the distribution p. Similarly, let $V_p[g]$ denote the variance of the function g at p.

Theorem 3. *For any $g : X \to \mathbb{R} \in \mathbb{R}^X$,*

$$V_p = ||(dE[g])_p||_p^2 = ||(\nabla E[g])_p||_p^2,$$

where the norm is induced by the Fisher information metric, for all $p \in P(X)$.

---

[1] Different coordinates are sometimes chosen in information geometry for $P(X)$, letting (for instance) $x_{n+1} = 1 - \sum_{i=1}^{n} x_i$ to enforce $\sum_i x_i = 1$. This yields an asymmetric set of replicator equations, so a different set of coordinates is chosen in this exposition.

3.3. **Information Divergences and Metrics on $P(X)$.** Some Riemannian metrics on $P(X)$ can be derived from information divergences [5]. Information geometry defines an *information divergence* as a smooth function $D(\cdot||\cdot) : P(X) \times P(X)$ such that $D(x||y) \geq 0$ with equality iff $x = y$. The second order Taylor expansion in either variable evaluated along the diagonal $x = y$ begins with the Hessian term H. Indeed,

$$D(x||y) = D(x||y)|_{x=y} + (\nabla D(x||y)|_{x=y}) \cdot (x - y) + \frac{1}{2} \cdot (x - y)^T H(x) \cdot (x - y) + \cdots$$

$$= 0 + 0 + \frac{1}{2} \cdot (x - y)^T H(x) \cdot (x - y) + \cdots$$

because the gradient is parallel to $\mathbf{1}$ and $\mathbf{1} \cdot (x - y) = \mathbf{1} \cdot x - \mathbf{1} \cdot y = 1 - 1 = 0$.

In the case that the Hessian is positive definite, it can be used to define a metric,

$$g_{ij}^{(D)} = \left(\frac{\partial^2 D}{\partial x_i \partial y_j} \right) |_{x=y}.$$

A metric then defines a gradient flow, hence a global information divergence yields a dynamical system on the simplex. Importantly, the Hessian of the Kullback-Liebler divergence (in either variable, evaluated on the diagonal) is the Fisher information matrix, yielding the local to global connection of these two measures of information.

Example 4 (Kullback-Liebler Divergence)**.** The Kullback-Liebler divergence localizes to the Fisher information metric. In coordinates we obtain the Shahshahani metric since

$$\frac{\partial^2}{\partial x_i \partial y_j} D_{KL}(x||y)|_{x=y} = \frac{1}{x_i} \delta_{ij}.$$

Hence the induced gradient flow is the replicator equation. This allows the interpretation of Fisher's Fundamental theorem and Kimura's maximal principle in terms of Fisher information: natural selection forms a gradient with respect to an informatic measure, and hence locally has the direction of maximal information increase. The rate of change of the mean fitness of the population is given by the informatic variance.

3.4. **Kullback-Liebler Divergence is a Lyapunov function for the Replicator Dynamic.** The following theorem shows that the Kullback-Liebler information divergence forms a Lyapunov function for the replicator dynamic, given an evolutionarily stable state. In fact, evolutionary stability is characterized by this property. A version of this theorem was proved in [1] and in [3].

Theorem 5. *The state $\hat{x}$ is an interior ESS for the replicator dynamic if and only if $D_{KL}(\hat{x}||x)$ is a local Lyapunov function.*

Proof. Let $V(x) = D_{KL}(\hat{x}||x) = \sum_i \hat{x}_i \log \hat{x}_i - \sum_i \hat{x}_i \log x_i$. Then we have that

$$\dot{V}(x) = -\sum_i \hat{x}_i \frac{\dot{x}_i}{x_i} = -\sum_i \hat{x}_i(f_i(x) - \bar{f}(x))$$

$$= -\sum_i \hat{x}_i f_i(x) + \sum_i \hat{x}_i \bar{f}(x) = -\sum_i \hat{x}_i f_i(x) + \left(\sum_i \hat{x}_i\right)\bar{f}(x)$$

$$= -\sum_i \hat{x}_i f_i(x) + \bar{f}(x) = -(\hat{x} \cdot f(x) - x \cdot f(x)) < 0.$$

The last inequality holds if and only if $\hat{x}$ is an ESS. Finally, by Jensen's inequality, D_{KL} is minimized when $x = \hat{x}$, so it is a local Lyapunov function. $\qquad\square$

A similar result is proven in [8], with the Lyapunov function $V(x) = \prod_i x_i^{\hat{x}_i}$, but the informatic origin is not apparent in this form, although the quantity V can be interpreted as the probability of finding a categorical distribution on x in the state $\hat{x}$. The quantity $D_{KL}(\hat{x}||x)$ can be described as the *potential information* of the replicator system. The above result can then be interpreted information theoretically – natural selection acts to minimize the potential information.

Theorem 5 holds for a class of ecological dynamics. A dynamic of the form $\dot{x}_i = x_i g_i(x)$, $i = 1, \ldots, n$ (an *ecological dynamic*) is called *aggregate monotone* with respect to a fitness landscape f if $g = (g_1, \ldots, g_n)$ has the property that $y \cdot f(x) > z \cdot f(x)$ if and only if $y \cdot g(x) > z \cdot g(x)$, for all distributions x, y, z. An aggregate monotone dynamic is the replicator dynamic up to a change in velocity [11]. In particular, the replicator equation with a convex function applied to the fitness landscape is aggregate monotone. Consider the following extension of Theorem 5.

Theorem 6. *For an aggregate monotone ecological dynamic $\dot{x}_i = x_i g_i(x)$, $D_{KL}(\hat{x}||x)$ is a Lyapunov function for the dynamic if $\hat{x}$ is an interior ESS.*

Proof. Let $V(x) = D_{KL}(\hat{x}||x) = \sum_i \hat{x}_i \log \hat{x}_i - \sum_i \hat{x}_i \log x_i$. Note that since $\dot{x}_i = x_i g_i(x)$ is a dynamic on the simplex, $0 = \sum_i x_i g_i(x) = x \cdot g(x)$. Then we have that

$$\dot{V}(x) = -\sum_i \hat{x}_i \frac{\dot{x}_i}{x_i} = -\sum_i \hat{x}_i g_i(x)$$

$$= -\hat{x} \cdot g(x) = -(\hat{x} \cdot g(x) - x \cdot g(x))$$

Applying aggregate monotonicity to the last equality completes the proof. $\qquad\square$

Since a change of velocity does not alter the orbits of the dynamic, Theorem 6 shows that the replicator equation is essentially the only aggregate monotone ecological dynamic in which evolutionary stability corresponds to minimizing the Kullback-Liebler divergence. For exactly which class of evolutionary dynamics this property holds for is an open question. From the proof it is clear that the assumption of aggregate monotonicity is too strong for a full characterization since it is only needed that $\hat{x} \cdot g(x) - x \cdot g(x) > 0$ if $\hat{x} \cdot f(x) - x \cdot f(x) > 0$, which quantifies over two distributions rather than three.

3.5. Exponential Families as Solutions of the Replicator Equation. The exponential map on the Shahshahani manifold is

$$exp(x, v) = \sum_i \frac{x_i e^{v_i}}{\sum_j x_j e^{v_j}} \hat{e}_i,$$

where $\hat{e}_i$ is the i-th coordinate vector [6]. The exponential map reduces to the exponential family at the barycenter $b = (\frac{1}{n}, \ldots, \frac{1}{n})$,

$$exp(b, v) = \sum_i \frac{\frac{1}{n} e^{v_i}}{\sum_j \frac{1}{n} e^{v_j}} \hat{e}_i = \frac{1}{\sum_j e^{v_j}} (e^{v_1}, \ldots, e^{v_n}).$$

The solutions of the replicator equation can be realized as exponential families [2, 6, 9]. Let $x_i = \exp(v_i - G)$ with $\dot{v}_i = f_i(x)$ and $G(x)$ a normalization constant to ensure that the distribution sums to one. From the fact that $\sum_i x_i = 1$, $0 = \sum_i \dot{x}_i$ and so

$$0 = \sum_i \dot{x}_i = \sum_i \exp(v_i(x) - G(x))(\dot{v}_i(x) - \dot{G}(x))$$

$$= \sum_i x_i(\dot{v}_i(x) - \dot{G}(x)) = \sum_i (x_i f_i(x)) - \dot{G}(x)$$

$$= \bar{f}(x) - \dot{G}(x)$$

Hence $\dot{G} = \bar{f}(x)$. Now x_i satisfies

$$\dot{x}_i = \exp(v_i(x) - G(x))(\dot{v}_i(x) - \dot{G}(x)) = x_i(f_i(x) - \bar{f}(x)),$$

which is the replicator equation. In the case of a log-linear fitness landscape, explicit solutions can be derived [6]. In this case, the equation for the variable v can be reduced to a linear differential equation, which can be solved with eigenvalue methods.

3.6. Denormalization.

Information geometry defines the *denormalized manifold* $\tilde{P}(X) = \{\tau p | \tau \in \mathbb{R}^+, p \in P(X)\}$, which can be thought of as non-normalized discrete probability distributions. As with $P(X)$, $\tilde{P}(X)$ has an information metric. The denormalized manifold embeds into the reals as $\mathbb{R}^n_+$, with the denormalized information metric induced by the the metric given by Shahshahani, where the mapping back onto $P(X)$ realizes τ as $|x|$. In coordinates, the metric is given by

$$\tilde{g}_{ij}(x) = \tau g_{ij}(x) = \frac{\tau}{x_i}\delta_{ij},$$

Akin uses the metric

$$g_{ij}(x) = \frac{1}{x_i}\delta_{ij}$$

on $\mathbb{R}^n_+$ rather than the metric

$$g_{ij}(x) = \frac{|x|}{x_i}\delta_{ij}$$

given by Shahshahani [2, 13]. Both metrics restrict to the same metric given by Shahshahani on the simplex. From the point of view of information geometry, the metric given by Shahshahani is the natural choice. The choice affects the form of the gradient on $\mathbb{R}^n_+$, which is in the case of Akin's metric is the Lotka-Volterra predator-prey equation.

3.7. The Lotka-Volterra Equations and the Replicator Equation.

The Lotka-Volterra equations

$$(1) \qquad\qquad\qquad \dot{x}_i = x_i f_i(x)$$

descend from $\mathbb{R}^n_+$, through a normalization map onto the simplex, to a replicator equation with an altered landscape. To see this, let $|x| = x_1 + \cdots + x_n$, $\dot{x}_i = x_i f_i(x)$ and $y_i = \frac{x_i}{|x|}$. Rearrange to $|x|y_i = x_i$ and note that $\frac{d}{dt}|x| = \sum_i \dot{x}_i = \sum_i x_i f_i(x) = x \cdot f(x)$. By the product rule, $\frac{d}{dt}|x|y_i + |x|\dot{y}_i = \dot{x}_i$ and so

$$
\begin{aligned}
\frac{d}{dt}y_i &= \frac{\dot{x}_i - \frac{d}{dt}|x|\dot{y}_i}{|x|} \\
&= \frac{x_i f_i(x) - x \cdot f(x)y_i}{|x|} \\
&= y_i(f_i(x) - y \cdot f(x)) \\
&= y_i(g_i(y) - y \cdot g(y)),
\end{aligned}
$$

where $g_i(y) = f_i(x)$ is an alteration of the fitness landscape.

The Lotka-Volterra equations are the gradient flow with respect to the metric given by Akin on $\mathbb{R}^n_+$ [8]. The gradient of the metric given by Shahshahani differs by a factor of $|x|$:

$$\dot{x}_i = \frac{x_i}{|x|} f_i(x), \tag{2}$$

This system is transformable to Equation 1 after a change of velocity eliminating the scalar function $B(x) = \frac{1}{|x|}$ because B is strictly positive on $\mathbb{R}^n_+$. Equation 2 transforms to a replicator equation via the normalization map [13].

The Lotka-Volterra equations can be interpreted as a gradient of the denormalized Fisher information metric, in the case that f is an Euclidean metric, in analogy to the replicator equation. This allows denormalized analogues of earlier results, such as the following, which is true for the denormalized version of the Lotka-Volterra equation.

Theorem 7. *Let $\hat{x}$ in $\mathbb{R}^n_+$ be such that*

$$\frac{\hat{x} \cdot f(x)}{|\hat{x}|} > \frac{x \cdot f(x)}{|x|}$$

in some neighborhood of $\hat{x}$ (a denormalized ESS). Suppose that the trajectory of $\dot{x}_i = \frac{x_i}{|x|} f_i(x)$ lies in a set that contains no point parallel to $\hat{x}$. Then the denormalized Kullback-Liebler divergence

$$D_{KL}\left(\frac{\hat{x}}{|\hat{x}|} || \frac{x}{|x|} \right)$$

is a local Lyapunov function for Equation 2.

Proof. The divergence is minimal (and equal to zero) when $x = c\hat{x}$ for some constant c. Hence if the line through the origin and the point $\hat{x}$ intersects the the trajectory at most once, the divergence is zero if and

only if $\hat{x} = x$. The time derivative is

$$\frac{d}{dt}\left[D_{KL}\left(\frac{\hat{x}}{|\hat{x}|}||\frac{x}{|x|}\right)\right] = 0 - \frac{d}{dt}\left[\sum_i \frac{\hat{x}_i}{|\hat{x}|}(\log x_i - \log|x|)\right]$$

$$= -\sum_i \frac{\hat{x}_i}{|\hat{x}|}\frac{\dot{x}_i}{x_i} + \sum_i \frac{\hat{x}_i}{|\hat{x}|}\frac{\dot{|x|}}{|x|}$$

$$= -\sum_i \frac{\hat{x}_i}{|\hat{x}|}\frac{1}{x_i}\frac{x_i}{|x|}f_i(x) + \frac{x \cdot f(x)}{|x|^2}$$

$$= -\frac{1}{|x|}\sum_i \frac{\hat{x}_i}{|\hat{x}|}f_i(x) + \frac{1}{|x|^2}x \cdot f(x)$$

$$= -\frac{1}{|x|}\left(\frac{\hat{x} \cdot f(x)}{|\hat{x}|} - \frac{x \cdot f(x)}{|x|}\right) < 0.$$

$\square$

4. Informatics of Multiple Population Replicator Dynamics

The information-theoretic approach easily extends to multiple population replicator equations such as bimatrix games. As before, the potential information plays a crucial role. It is the sum of the potential informations of all populations that plays the role of the Lyapunov function and gives rise to the geometry. It suffices to discuss the two population case as it is clear that the results extend inductively to finitely-many populations.

4.1. **Two Populations.** Consider two categorical distributions $p = (p_1, \ldots, p_n)$ and $q = (q_{n+1}, \ldots, q_{n+m})$ with fitness landscapes $f(p, q) = (f_1(p, q), \ldots, f_n(p, q))$ and $g(p, q) = (g_{n+1}(p, q), \ldots, g_{n+m}(p, q))$. Define the coupled replicator system

$$\dot{p}_i = p_i(f_i(p, q) - \mathbb{E}_p[f(p, q)])$$
$$\dot{q}_j = q_j(g_j(p, q) - \mathbb{E}_q[g(p, q)])$$

where i runs from 1 to n and j runs from $n+1$ to $n+m$. Note carefully that the expected values are taken with each distribution respectively.

This system is the gradient flow of the Riemannian metric defined on the interior of $\Delta^n \times \Delta^m$ given by

$$G_{i,j}(p,q) = \begin{cases} \frac{1}{p_i} & \text{if } i = j \le n \\ \frac{1}{q_i} & \text{if } i = j > n \\ 0 & \text{else} \end{cases}$$

That is, the matrix is the direct sum matrix of the usual metric for each equation. As in the single population case, we can use potential information to form a Lyapunov function for the system. Given states $\hat{p}$ and $\hat{q}$, let L be the sum of the potential information of each categorical distribution. That is,

$$L = D_p(\hat{p}, p) + D_q(\hat{q}, q)$$
$$= \sum_i \hat{p}_i \log \hat{p}_i - \sum_i \hat{p}_i \log p_i + \sum_j \hat{q}_j \log \hat{q}_j - \sum_j \hat{q}_j \log q_j$$

The metric can be obtained as the localization of the sum of the divergence functions. All the usual calculations follow from the fact that the system is a gradient, e.g. Fisher's Fundamental Theorem.

4.2. **Potential Information is a Lyapunov Function.** Recall that $\hat{p}$ is an ESS in the single population case if $\hat{p} \cdot f(p) > p \cdot f(p)$ for all p in a neighborhood of $\hat{p}$.

Theorem 8. *If $\hat{p}$ and $\hat{q}$ are ESS for each system respectively then L is a Lyapunov function for the coupled system.*

Proof. A straight-forward computation shows that (up to a negative)

$$\dot{L} = \hat{p} \cdot f(p,q) - p \cdot f(p,q) + \hat{q} \cdot g(p,q) - q \cdot g(p,q).$$

L is positive everywhere and has minimum at $(\hat{p}, \hat{q})$. Since $\hat{p}$ and $\hat{q}$ are ESS, $\dot{L} < 0$, so L is a local Lyapunov function.   □

Notice that the hypothesis that both $\hat{p}$ and $\hat{q}$ are ESS is too strong. Indeed, all that is required is that

$$\hat{p} \cdot f(p,q) + \hat{q} \cdot g(p,q) > p \cdot f(p,q) + q \cdot g(p,q).$$

Call this condition a *coupled ESS* (as well as its obvious higher dimensional analogs) and note that any ESS is a coupled ESS. Then L is a Lyapunov for the system if and only if ($\hat{p}$ and $\hat{q}$) is a coupled ESS for the two population system.

4.3. **Solutions.** We can again show that the solutions are exponential families. Let $\dot{v} = f(p, q)$ and $\dot{w} = g(p, q)$. Let $\dot{N} = \mathbb{E}_p\left[f(p, q)\right]$ and $\dot{M} = \mathbb{E}_q\left[g(p, q)\right]$. Then $p_i = \exp\left(v_i - N\right)$ and $q_j = \exp\left(w_j - M\right)$ is a solution to the coupled system. Indeed,

$$\dot{p}_i = \exp\left(v_i - N\right)(\dot{v}_i - \dot{N}) = p_i(f_i(p, q) - \mathbb{E}_p\left[f(p, q)\right]),$$

and similarly for q_j.

4.4. **Multiple Populations.** The above generalizes by induction to show that for a coupled system of multiple interacting populations, the sum of the respective potential informations gives a Lyapunov function for a coupled ESS.

5. Discussion

The Shahshahani geometry can be interpreted within the framework of information theory as the information geometry of the simplex. This explains the origin of several quantities in evolutionary game theory including the Shahshahani metric and the Kullback-Liebler information divergence. An important feature of the approach is that the information-geometric reasoning extends to the Lotka-Volterra equation and the multiple population replicator equation easily within the framework. Additionally, the replicator dynamic arises intuitively from purely mathematical and statistical concepts such as Fisher information. This shows that the replicator equation models the information dynamics of natural selection.

References

[1] Ethan Akin. The geometry of population genetics. *Lecture Notes in Biomathematics*, 17(30-31), 1979.
[2] Ethan Akin. Exponential families and game dynamics. *Canadian Journal of Mathematics*, XXXIV(2):374–405, 1982.
[3] Ethan Akin. The differential geometry of population genetics and evolutionary games. *Mathematical and Statistical Developments of Evolutionary Theory*, 299:1–94, 1990.
[4] Shunichi Amari and A Fujiwara. Gradient systems in view of information geometry. *Physica D: Nonlinear Phenomena*, 1995.
[5] Shunichi Amari and Hiroshi Nagaoka. *Methods of Information Geometry*, volume 191 of *Translations of Mathematical Monographs*. Oxford University Press, 1993.
[6] Nihat Ay and Ionas Erb. On a notion of linear replicator equations. *Journal of Dynamics and Differential Equations*, 17(2):427–451, 2005.
[7] N.N. Chentsov. Statistical decision rules and optimal inference. 1982.
[8] Josef Hofbauer and Karl Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[9] Georgiy P Karev. Replicator equations and the principle of minimal production of information. *ArXiv*.

[10] Martin Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Belknap Press of Harvard University Press, 2006.

[11] K. Ritzberger and J.W. Weibull. Evolutionary selection in normal form games. *Econometrica*, 63:1371–1399, 1995.

[12] William H. Sandholm. *Population Games and Evolutionary Dynamics*. Cambridge, 2008.

[13] S. Shahshahani. A new mathematical framework for the study of linkage and selection. *Memoirs of the AMS*, 17(221), 1979.

UNIVERSITY OF CALIFORNIA LOS ANGELES
E-mail address: `marcharper@ucla.edu`