Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Linear Algebra and its Applications 396 (2005) 259–272
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

A Perron Theorem for positive componentwise bilinear maps

Joseph E. Carroll *, Timothy Lauck, Roland H. Lamberson

Department of Mathematics, Humboldt State University, Arcata, CA 95501, United States

Received 18 December 2003; accepted 17 September 2004

Submitted by R.A. Brualdi

Abstract

It is proved that a Perron type theorem holds for positive maps with bilinear components whose defining matrices satisfy a maximality assumption with respect to certain entry ratios. The result is applied to a life history model which includes sexual reproduction.

AMS classification: 15A48; 92D25

Keywords: Bilinear map; Contraction mapping; Hilbert projective pseudometric; Perron–Frobenius theorem

1. Introduction

The common proof of the Perron–Frobenius Theorem, which gives insights into the behavior of the successive iterates of any positive square matrix as a function on the positive real cone of corresponding dimension, is an algebraic one. However, a large part of this theorem can be readily proved analytically using the theorem of Birkhoff, [1] and [2–pp. 383–385], which states that such a matrix induces a contraction mapping on the projective quotient of the cone with respect to the Hilbert

* Corresponding author.
E-mail address: jcrrll@cox.net (J.E. Carroll).

0024-3795/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2004.09.021

projective metric. With coauthors, Kohlberg [3,4–and unpublished work with Pratt] has further investigated this metric. In addition, Kohlberg [5] and other investigators, e.g., [6], have established that some functions on the positive cone which share certain properties with positive matrices, including nonnegativity and homogeneity of degree one, are contraction mappings in the same sense and, therefore, have associated Perron theorems. Here, we study positive mappings defined on the product of two positive cones with bilinear vector component functions, and their compositions with some homogeneous maps on each of the two vector coordinates separately. As we shall see, such mappings can be useful in the description of life histories of some simple populations which reproduce sexually.

2. The basic bilinear model

The motivation for this paper is a reproductive model. We consider a two sex population with m (pheno)types of females and n types of males, and assume type inheritance as follows. For $i = 1, \dots, m$ and $j = 1, \dots, n$, let for $k = 1, \dots, m$, $\sigma_{kij} \geqslant 0$ be the proportion of female offspring from a mating between a type i female and a type j male which are of type k and, for $k = 1, \dots, n$, $\tau_{kij} \geqslant 0$ be the proportion of male offspring from the same mating which are of type k. Phenotypic inheritance of this sort might occur for traits determined by multiple genes or might be simply all that is observable if the genotypes which determine the phenotypes are unknown. Clearly, for any (i, j), $\sum_{k=1}^{m} \sigma_{kij} = 1 = \sum_{k=1}^{n} \tau_{kij}$. For $k = 1, \dots, m$, let $\mathbf{S}_k$ be the $m \times n$ matrix whose (i, j)th term is σ_{kij} and for $k = 1, \dots, n$, let $\mathbf{T}_k$ be the $m \times n$ matrix whose (i, j)th term is τ_{kij}. Then $\sum_{k=1}^{m} \mathbf{S}_k = \sum_{k=1}^{n} \mathbf{T}_k$ is the constant $m \times n$ matrix with all entries 1.

For any positive integer m, let $\mathbf{R}_+^m = \{(x_1, \dots, x_m) \in \mathbf{R}^m - \{\mathbf{0}\} \mid x_i \geqslant 0$ for $i = 1, \dots, m\}$, the nonnegative cone in $\mathbf{R}^m$. Let $|\cdot|$ denote the $\boldsymbol{\ell}_1$ norm on $\mathbf{R}^m$, i.e., $|\mathbf{x}| = |x_1| + \cdots + |x_m|$, and observe that for $\mathbf{x} \in \mathbf{R}_+^m$, $|\mathbf{x}| = x_1 + \cdots + x_m$. Let $\mathbf{H}_+^{m-1} = \{\mathbf{x} \in \mathbf{R}_+^m \mid |\mathbf{x}| = 1\}$, the space of nonnegative stochastic vectors in $\mathbf{R}^m$, a compact subset of an $m - 1$ dimensional hyperplane. Initially, we shall track the phenotypic proportions, rather than the absolute numbers, of the female and male populations. Then, female vectors lie in $\mathbf{H}_+^{m-1}$ and male vectors in $\mathbf{H}_+^{n-1}$, so the space of population vectors is $\mathscr{H}_+ = \mathbf{H}_+^{m-1} \times \mathbf{H}_+^{n-1}$. Mating is assumed to be random and the offspring become the next generation. Therefore, if $\mathbf{x}(t)$ represents the female vector of the tth generation and $\mathbf{y}(t)$ the male vector, the transformation on $\mathscr{H}_+$ which carries one generation to the next uses the $\mathbf{S}_k$ and $\mathbf{T}_k$ as bilinear forms, namely

$$x_k(t + 1) = \mathbf{x}(t)^{\mathrm{T}} \mathbf{S}_k \mathbf{y}(t), \quad k = 1, \dots, m, \tag{1a}$$

$$y_k(t + 1) = \mathbf{x}(t)^{\mathrm{T}} \mathbf{T}_k \mathbf{y}(t), \quad k = 1, \dots, n. \tag{1b}$$

The summation conditions on the $\mathbf{S}_k$'s and $\mathbf{T}_k$'s imply that $(\mathbf{x}(t + 1), \mathbf{y}(t + 1)) \in \mathscr{H}_+$, as we wish, so our model is that of a discrete dynamical system on $\mathscr{H}_+$. Let us define bilinear maps $B_1 : \mathscr{H}_+ \to \mathbf{H}_+^{m-1}$ and $B_2 : \mathscr{H}_+ \to \mathbf{H}_+^{n-1}$ by $B_1(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T\mathbf{S}_1\mathbf{y}, \ldots, \mathbf{x}^T\mathbf{S}_m\mathbf{y})$ and $B_2(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T\mathbf{T}_1\mathbf{y}, \ldots, \mathbf{x}^T\mathbf{T}_n\mathbf{y})$. Then let $\boldsymbol{B} : \mathscr{H}_+ \to \mathscr{H}_+$ be defined by $\boldsymbol{B}(\mathbf{x}, \mathbf{y}) = (B_1(\mathbf{x}, \mathbf{y}),\ B_2(\mathbf{x}, \mathbf{y}))$. Eqs. (1) state that $(\mathbf{x}(t+1), \mathbf{y}(t+1)) = \boldsymbol{B}(\mathbf{x}(t), \mathbf{y}(t))$ for $t \geqslant 0$, so $(\mathbf{x}(t), \mathbf{y}(t)) = \boldsymbol{B}^t(\mathbf{x}(0), \mathbf{y}(0))$.

3. Hardy–Weinberg equilibrium

There is a classic instance of this model. One of the simplest sets of phenotypes is that of the three genotypes associated with a diploid autosomal gene locus with two codominant alleles, A and B. The phenotypes are AA, AB, and BB, which we number 1, 2, and 3 respectively, for both females and males, and it is easily calculated that

$$\mathbf{S}_1 = \mathbf{T}_1 = \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{S}_2 = \mathbf{T}_2 = \begin{pmatrix} 0 & \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & \frac{1}{2} & 0 \end{pmatrix},$$

$$\mathbf{S}_3 = \mathbf{T}_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{pmatrix}.$$

The Hardy–Weinberg Law states that if $\mathbf{x}(0) = (x_1, x_2, x_3) = (y_1, y_2, y_3) = \mathbf{y}(0)$, then for all $t \geqslant 1, \mathbf{x}(t) = \mathbf{y}(t) = (p^2, 2pq, q^2)$, where $p = x_1 + \frac{1}{2}x_2$ and $q = \frac{1}{2}x_2 + x_3$ are the allele frequencies for A and B. This is easily verified by directly calculating $(\mathbf{x}(1), \mathbf{y}(1)) = \boldsymbol{B}(\mathbf{x}(0), \mathbf{y}(0))$, and then noting that the allele frequencies for $(p^2, 2pq, q^2)$ are again p and q. In fact, even if $\mathbf{x}(0) \neq \mathbf{y}(0)$, the vectors of phenotype proportions stabilize in two generations since the fact that $\mathbf{S}_k = \mathbf{T}_k$ for $k = 1, 2, 3$ implies that $\mathbf{x}(1) = \mathbf{y}(1)$. Therefore, $(\mathbf{x}(t), \mathbf{y}(t))$ is constant for $t \geqslant 2$, but its value depends on the initial condition, $(\mathbf{x}(0), \mathbf{y}(0)) \in \mathscr{H}_+$. Our main results will be of a different sort and will not apply in cases, like this one, where any of the $\mathbf{S}_k$'s or $\mathbf{T}_k$'s has a zero entry.

4. The Hilbert projective pseudometric and Birkhoff's theorem

Let $m \geqslant 2$ and let $\mathbf{R}_{++}^m = \{(x_1, \ldots, x_m) \in \mathbf{R}^m \mid x_i > 0 \text{ for } i = 1, \ldots, m\}$, an abelian group under coordinatewise multiplication, with identity $\mathbf{1} = (1, \ldots, 1)$. We define a norm, $\|\| : \mathbf{R}_{++}^m \to [1, \infty)$, which will be manipulated multiplicatively rather than additively, by $\|\mathbf{x}\| = \frac{\max_{1 \leqslant i \leqslant m} x_i}{\min_{1 \leqslant i \leqslant m} x_i} = \max_{1 \leqslant i, j \leqslant m} \frac{x_i}{x_j}$. One can easily verify the following four observations, the first three of which are analogous to additive properties of real vector space norms. For all $\mathbf{x},\ \mathbf{x}' \in \mathbf{R}_{++}^m,\ \|\mathbf{x}\| = 1$ if and only $\mathbf{x}$ is constant or scalar, i.e., $\mathbf{x} = (x, \ldots, x)$ for some $x > 0$; $\|\mathbf{x}\| = \|\mathbf{x}^{-1}\|$; $\|\mathbf{x}\mathbf{x}'\| \leqslant \|\mathbf{x}\|\|\mathbf{x}'\|$; for all $c > 0$, $\|c\mathbf{x}\| = \|\mathbf{x}\|$, i.e., $\|\|$ is homogeneous of degree zero. Using the group structure, we define a distance d on $\mathbf{R}^m_{++}$, again in analogy to the vector space case, by $d(\mathbf{x}, \mathbf{x}') = \|\mathbf{x}\mathbf{x}^{-1}\|$. By the norm properties, (i) $d(\mathbf{x}, \mathbf{x}') \geqslant 1$ and $d(\mathbf{x}, \mathbf{x}') = 1$ if and only if $\mathbf{x} = c\mathbf{x}'$ for some $c > 0$; (ii) $d(\mathbf{x}, \mathbf{x}') = d(\mathbf{x}', \mathbf{x})$; (iii) $d(\mathbf{x}, \mathbf{x}'') \leqslant d(\mathbf{x}, \mathbf{x}')d(\mathbf{x}', \mathbf{x}'')$; (iv) $d(a\mathbf{x}, b\mathbf{x}') = d(\mathbf{x}, \mathbf{x}')$ for all $a, b > 0$. The Hilbert projective pseudometric δ on $\mathbf{R}^m_{++}$ is defined by $\delta(\mathbf{x}, \mathbf{x}') = \log d(\mathbf{x}, \mathbf{x}')$. He introduced it in 1903 [7], applying it to Bolyai-Lobachevsky geometry. By (i), (ii), and (iii), it satisfies the definition of a metric except that $\delta(\mathbf{x}, \mathbf{x}') = 0$ if and only if $\mathbf{x} = c\mathbf{x}'$ for some $c > 0$.

Let $\mathbf{P}^{m-1}$ denote real $m-1$ dimensional projective space, let $\mathbf{P}^{m-1}_{+} = \{\mathbf{x} \in \mathbf{P}^{m-1} \mid \mathbf{x}$ has projective coordinates in $\mathbf{R}^m_{+}\}$, and let $\mathbf{P}^{m-1}_{++} = \{\mathbf{x} \in \mathbf{P}^{m-1} \mid \mathbf{x}$ has projective coordinates in$\mathbf{R}^m_{++}\}$. We observe that properties (i) and (iv) of d above imply that δ induces metrics both on $\mathbf{P}^{m-1}_{++}$ and on $\mathbf{H}^{m-1}_{++} = \mathbf{H}^{m-1}_{+} \cap \mathbf{R}^m_{++}$. We shall return to the former, and more important of these, later.

We can define a distance on $\mathscr{H}_{++} = \mathbf{H}^{m-1}_{++} \times \mathbf{H}^{n-1}_{++}$ which we again call d, by $d((\mathbf{x}, \mathbf{y}), (\mathbf{x}', \mathbf{y}')) = d(\mathbf{x}, \mathbf{x}')d(\mathbf{y}, \mathbf{y}')$, and the corresponding Hilbert metric δ by $\delta((\mathbf{x}, \mathbf{y}), (\mathbf{x}', \mathbf{y}')) = \log d((\mathbf{x}, \mathbf{y}), (\mathbf{x}', \mathbf{y}')) = \delta(\mathbf{x}, \mathbf{x}') + \delta(\mathbf{y}, \mathbf{y}')$. Our goal is to prove that when certain conditions on the $\mathbf{S}_k$ and $\mathbf{T}_k$ are satisfied, then $\boldsymbol{B}$ is a contraction mapping on $\mathscr{H}_{++}$ with respect to this δ. We shall require the following result, which is part of a theorem due to Birkhoff. Proofs may be found in [2–pp. 383–385], [8–pp. 100–110], [9], and, with this notation, in [10].

Theorem 1 (Birkhoff). *Let $k, l \geqslant 2$ and let $\mathbf{A}$ be an $k \times l$ matrix with positive entries, so that $\mathbf{A} : \mathbf{R}^l_{+} \to \mathbf{R}^k_{++}$. For $i = 1, \ldots, k$, let $\mathbf{a}_i$ be the ith row vector of $\mathbf{A}$, and let $d(\mathbf{A}) = \max_{1 \leqslant i, j \leqslant k} d(\mathbf{a}_i, \mathbf{a}_j)$. Then for any $\mathbf{x}, \mathbf{x}' \in \mathbf{R}^l_{++}$, $\delta(\mathbf{x}, \mathbf{x}') = 0$ or $d(\mathbf{A}) = 1$ implies that $\delta(\mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}') = 0$. Otherwise, $\delta(\mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}') < T(\mathbf{A})\delta(\mathbf{x}, \mathbf{x}')$, where $T(\mathbf{A}) = \frac{\sqrt{d(\mathbf{A})}-1}{\sqrt{d(\mathbf{A})}+1}$ is minimal with this property.*

5. Iterated bilinear maps

Armed with the machinery of the last section, we return to the model of Section 2, but we do not require the summation conditions on the $\mathbf{S}_k$'s and $\mathbf{T}_k$'s for our arguments. After noting that for any $(\mathbf{x}, \mathbf{y}) \in \mathscr{R}_{+} = \mathbf{R}^m_{+} \times \mathbf{R}^n_{+}$, the 1×1 matrix $\mathbf{x}^{\mathrm{T}}\mathbf{S}_k\mathbf{y} = (\mathbf{x}^{\mathrm{T}}\mathbf{S}_k\mathbf{y})^{\mathrm{T}} = \mathbf{y}^{\mathrm{T}}\mathbf{S}_k^{\mathrm{T}}\mathbf{x}$ and, similarly, $\mathbf{x}^{\mathrm{T}}\mathbf{T}_k\mathbf{y} = (\mathbf{x}^{\mathrm{T}}\mathbf{T}_k\mathbf{y})^{\mathrm{T}} = \mathbf{y}^{\mathrm{T}}\mathbf{T}_k^{\mathrm{T}}\mathbf{x}$, we define matrices $\mathbf{B}_1(\mathbf{x}) \in \mathbf{R}^{m \times n}_{+}$, $\mathbf{B}_1(\mathbf{y}) \in \mathbf{R}^{m \times m}_{+}$, $\mathbf{B}_2(\mathbf{x}) \in \mathbf{R}^{n \times n}_{+}$, and $\mathbf{B}_2(\mathbf{y}) \in \mathbf{R}^{n \times m}_{+}$ and observe that

$$B_1(\mathbf{x}, \mathbf{y}) = \mathbf{B}_1(\mathbf{x})\mathbf{y} = \begin{pmatrix} \mathbf{x}^{\mathrm{T}}\mathbf{S}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}^{\mathrm{T}}\mathbf{S}_m \end{pmatrix} \mathbf{y} = \mathbf{B}_1(\mathbf{y})\mathbf{x} = \begin{pmatrix} \mathbf{y}^{\mathrm{T}}\mathbf{S}_1^{\mathrm{T}} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}^{\mathrm{T}}\mathbf{S}_m^{\mathrm{T}} \end{pmatrix} \mathbf{x}, \qquad (1a')$$

$$B_2(\mathbf{x},\mathbf{y}) = \mathbf{B}_2(\mathbf{x})\mathbf{y} = \begin{pmatrix} \mathbf{x}^{\mathrm{T}}\mathbf{T}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}^{\mathrm{T}}\mathbf{T}_n \end{pmatrix} \mathbf{y} = \mathbf{B}_2(\mathbf{y})\mathbf{x} = \begin{pmatrix} \mathbf{y}^{\mathrm{T}}\mathbf{T}_1^{\mathrm{T}} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}^{\mathrm{T}}\mathbf{T}_n^{\mathrm{T}} \end{pmatrix} \mathbf{x}. \tag{1b$'$}$$

If it happens to be the case that the $\mathbf{S}_k$'s and $\mathbf{T}_k$'s satisfy the summation conditions of Section 2 and $(\mathbf{x},\mathbf{y}) \in \mathscr{H}_+$, then $\mathbf{B}_1(\mathbf{x})$, $\mathbf{B}_1(\mathbf{y})$, $\mathbf{B}_2(\mathbf{x})$, and $\mathbf{B}_2(\mathbf{y})$ are all transpose stochastic, i.e., their column vectors lie in $\mathbf{H}_+^{m-1}$ or $\mathbf{H}_+^{n-1}$. The following two lemmas are found in any treatment of the Hilbert projective pseudometric.

Lemma 1. *Let* $m, n \geqslant 2$ *and* $\mathbf{S}$ *be a positive* $m \times n$ *matrix with rows* $\mathbf{r}_1, \ldots, \mathbf{r}_m$ *and columns* $\mathbf{c}_1, \ldots, \mathbf{c}_n$. *Then* $\max_{1\leqslant i,h\leqslant m} d(\mathbf{r}_i, \mathbf{r}_h) = \max_{1\leqslant j,l\leqslant n} d(\mathbf{c}_j, \mathbf{c}_l)$.

Proof. If $\mathbf{S} = (s_{ij})$, then both of these maxima are equal to $\max_{1\leqslant i,h\leqslant m; 1\leqslant j,l\leqslant n} \frac{s_{ij}s_{hl}}{s_{il}s_{hj}}$. □

Lemma 2. *For all* $\mathbf{c}, \mathbf{c}' \in \mathbf{R}_{++}^m$ *and* $\mathbf{x} \in \mathbf{R}_+^m$, $\frac{\mathbf{x}\cdot\mathbf{c}}{\mathbf{x}\cdot\mathbf{c}'} \leqslant \max_{1\leqslant i\leqslant m} \frac{c_i}{c_i'}$.

Proof

$$\frac{\mathbf{x}\cdot\mathbf{c}}{\mathbf{x}\cdot\mathbf{c}'} = \frac{\sum_{i=1}^m x_i c_i}{\sum_{i=1}^m x_i c_i'} = \sum_{i=1}^m \frac{x_i c_i' \frac{c_i}{c_i'}}{\sum_{i=1}^m x_i c_i'}$$

which is a weighted average of the terms $\frac{c_i}{c_i'}$, and so can be no greater than their maximum. □

Lemma 3. *Let* $\mathbf{S}$ *be as in Lemma* 1 *and let* $\mathbf{x} \in \mathbf{R}_+^m$. *Then* $\|\mathbf{x}^{\mathrm{T}}\mathbf{S}\| \leqslant \max_{1\leqslant i\leqslant m} \|\mathbf{r}_i\|$.

Proof. For some pair of columns of $\mathbf{S}$, say $\mathbf{c}$ and $\mathbf{c}'$, $\|\mathbf{x}^{\mathrm{T}}\mathbf{S}\| = \frac{\mathbf{x}\cdot\mathbf{c}}{\mathbf{x}\cdot\mathbf{c}'}$, and for each $i = 1, \ldots, m$, $\frac{c_i}{c_i'} \leqslant \|\mathbf{r}_i\|$. The result then follows from Lemma 2. □

Theorem 2. *Assume that all the* $\mathbf{S}_k$ *and* $\mathbf{T}_k$ *are positive matrices, not necessarily satisfying the summation conditions, but with* $\boldsymbol{B}$ *defined as in Section* 2. *Let* γ *be the maximum of all the row and column norms of all the* $\mathbf{S}_k$ *and* $\mathbf{T}_k$, *let* $\eta = \max\left\{\max_{i,j,k,l} \frac{\sigma_{kij}}{\sigma_{lij}}, \max_{i,j,k,l} \frac{\tau_{kij}}{\tau_{lij}}\right\}$, *and let* $\varphi = \min\{\gamma, \eta\}$. *Then, for any* $(\mathbf{x},\mathbf{y})$, $(\mathbf{x}',\mathbf{y}') \in \mathscr{R}_{++} = \mathbf{R}_{++}^m \times \mathbf{R}_{++}^n$, $\varphi = 1$ *or* $\delta((\mathbf{x},\mathbf{y}), (\mathbf{x}',\mathbf{y}')) = 0$ *implies that* $\delta(\boldsymbol{B}(\mathbf{x},\mathbf{y}), \boldsymbol{B}(\mathbf{x}',\mathbf{y}')) = 0$. *Otherwise*, $\delta(\boldsymbol{B}(\mathbf{x},\mathbf{y}), \boldsymbol{B}(\mathbf{x}',\mathbf{y}')) < 2\frac{\varphi-1}{\varphi+1}\delta((\mathbf{x},\mathbf{y}), (\mathbf{x}',\mathbf{y}'))$.

Proof. If $\gamma = 1$, then all the $\mathbf{S}_k$ and $\mathbf{T}_k$ are constant matrices, say $\sigma_{kij} = \sigma_k$ and $\tau_{kij} = \tau_k$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$. For any $(\mathbf{x},\mathbf{y}) \in \mathscr{R}_+$, $\mathbf{x}^{\mathrm{T}}\mathbf{S}_k\mathbf{y} = \sigma_k|\mathbf{x}\|\mathbf{y}|$, $\mathbf{x}^{\mathrm{T}}\mathbf{T}_k\mathbf{y} = \tau_k|\mathbf{x}\|\mathbf{y}|$. Therefore, $\boldsymbol{B}(\mathbf{x},\mathbf{y}) = |\mathbf{x}||\mathbf{y}|((\sigma_1, \ldots, \sigma_m), (\tau_1, \ldots, \tau_n))$, and so $\delta(\boldsymbol{B}(\mathbf{x},\mathbf{y}), \boldsymbol{B}(\mathbf{x}',\mathbf{y}')) = 0$ for any other $(\mathbf{x}',\mathbf{y}') \in \mathscr{R}_+$. If $\eta = 1$, then the $\mathbf{S}_k$

are all identical and so are the $\mathbf{T}_k$, so $B_1(\mathbf{x},\mathbf{y})$ and $B_2(\mathbf{x},\mathbf{y})$ are both constant vectors and, again, $\delta(\boldsymbol{B}(\mathbf{x},\mathbf{y}),\boldsymbol{B}(\mathbf{x}',\mathbf{y}'))=0$. Now suppose that $\varphi>1$ and let $(\mathbf{x},\mathbf{y})$, $(\mathbf{x}',\mathbf{y}')\in\mathscr{R}_{++}$. Then, by the triangle inequality,

$$\begin{aligned}\delta(B_1(\mathbf{x},\mathbf{y}),B_1(\mathbf{x}',\mathbf{y}'))&\leqslant\delta(B_1(\mathbf{x},\mathbf{y}),B_1(\mathbf{x}',\mathbf{y}))+\delta(B_1(\mathbf{x}',\mathbf{y}),B_1(\mathbf{x}',\mathbf{y}'))\\&=\delta(\mathbf{B}_1(\mathbf{y})\mathbf{x},\mathbf{B}_1(\mathbf{y})\mathbf{x}')+\delta(\mathbf{B}_1(\mathbf{x}')\mathbf{y},\mathbf{B}_1(\mathbf{x}')\mathbf{y}'),\end{aligned}$$

where this δ is defined on $\mathbf{R}^m_{++}$. Similarly, for δ defined on $\mathbf{R}^n_{++}$,

$$\delta(B_2(\mathbf{x},\mathbf{y}),B_2(\mathbf{x}',\mathbf{y}'))\leqslant\delta(\mathbf{B}_2(\mathbf{y})\mathbf{x},\mathbf{B}_2(\mathbf{y})\mathbf{x}')+\delta(\mathbf{B}_2(\mathbf{x}')\mathbf{y},\mathbf{B}_2(\mathbf{x}')\mathbf{y}')$$

so that for the appropriate δ's,

$$\begin{aligned}\delta(\boldsymbol{B}(\mathbf{x},\mathbf{y}),\boldsymbol{B}(\mathbf{x}',\mathbf{y}'))&=\delta(B_1(\mathbf{x},\mathbf{y}),B_1(\mathbf{x}',\mathbf{y}'))+\delta(B_2(\mathbf{x},\mathbf{y}),B_2(\mathbf{x}',\mathbf{y}'))\\&\leqslant\delta(\mathbf{B}_1(\mathbf{y})\mathbf{x},\mathbf{B}_1(\mathbf{y})\mathbf{x}')+\delta(\mathbf{B}_1(\mathbf{x}')\mathbf{y},\mathbf{B}_1(\mathbf{x}')\mathbf{y}')\\&\quad+\delta(\mathbf{B}_2(\mathbf{y})\mathbf{x},\mathbf{B}_2(\mathbf{y})\mathbf{x}')+\delta(\mathbf{B}_2(\mathbf{x}')\mathbf{y},\mathbf{B}_2(\mathbf{x}')\mathbf{y}').\end{aligned}$$

If $\delta((\mathbf{x},\mathbf{y}),(\mathbf{x}',\mathbf{y}'))=0$, then $\delta(\mathbf{x},\mathbf{x}')=\delta(\mathbf{y},\mathbf{y}')=0$ and, by Theorem 1, each of the four summands on the right above is 0, so $\delta(\boldsymbol{B}(\mathbf{x},\mathbf{y}),\boldsymbol{B}(\mathbf{x}',\mathbf{y}'))=0$. Suppose, then, that $\delta((\mathbf{x},\mathbf{y}),(\mathbf{x}',\mathbf{y}'))>0$. Again by Theorem 1, for the first of these summands,

$$\delta((\mathbf{B}_1(\mathbf{y})\mathbf{x},\mathbf{B}_1(\mathbf{y})\mathbf{x}')\leqslant T(\mathbf{B}_1(\mathbf{y}))\delta(\mathbf{x},\mathbf{x}')$$

with equality only if $\delta(\mathbf{x},\mathbf{x}')=0$ or $T(\mathbf{B}_1(\mathbf{y}))=0$.

By the definition of $\mathbf{B}_1(\mathbf{y})$ in Eq. (1a) and by Lemma 3, no row norm of $\mathbf{B}_1(\mathbf{y})$ exceeds γ, which implies that the maximum value of d on pairs of columns of $\mathbf{B}_1(\mathbf{y})$ does not exceed γ^2. Then by Lemma 1 and Theorem 1, $T(\mathbf{B}_1(\mathbf{y}))\leqslant\frac{\gamma-1}{\gamma+1}$, so $\delta(\mathbf{B}_1(\mathbf{y})\mathbf{x},\mathbf{B}_1(\mathbf{y})\mathbf{x}')\leqslant\frac{\gamma-1}{\gamma+1}\delta(\mathbf{x},\mathbf{x}')$ with equality only if $\delta(\mathbf{x},\mathbf{x}')=0$. After applying similar reasoning to the other three terms of the sum above and noting that at least one of $\delta(\mathbf{x},\mathbf{x}')$ and $\delta(\mathbf{y},\mathbf{y}')$ is positive, we obtain

$$\begin{aligned}\delta(\boldsymbol{B}(\mathbf{x},\mathbf{y}),\boldsymbol{B}(\mathbf{x}',\mathbf{y}'))&<2\frac{\gamma-1}{\gamma+1}\delta(\mathbf{x},\mathbf{x}')+2\frac{\gamma-1}{\gamma+1}\delta(\mathbf{y},\mathbf{y}')\\&=2\frac{\gamma-1}{\gamma+1}\delta((\mathbf{x},\mathbf{y}),(\mathbf{x}',\mathbf{y}')).\end{aligned}$$

Next, we reexamine $\mathbf{B}_1(\mathbf{y})$. For any distinct $k,l=1,\dots,m$, let $\mathbf{c}$ be the jth column vector of $\mathbf{S}_k^{\mathrm{T}}$ and $\mathbf{c}'$ the jth column vector of $\mathbf{S}_l^{\mathrm{T}}$. Applying Lemma 2, we conclude that $d(\mathbf{B}_1(\mathbf{y}))\leqslant\eta^2$, allowing us to proceed as above and eventually deduce that $\delta(\boldsymbol{B}(\mathbf{x},\mathbf{y}),\boldsymbol{B}(\mathbf{x}',\mathbf{y}'))<2\frac{\eta-1}{\eta+1}\delta((\mathbf{x},\mathbf{y}),(\mathbf{x}',\mathbf{y}'))$ as well. □

Next, we shall examine the behavior of the trajectory of $\boldsymbol{B}$. On $\mathbf{H}^{m-1}_{++}$ we shall be using both δ and Δ, the $\boldsymbol{\ell}_1$ metric defined by $\Delta(\mathbf{x},\mathbf{x}')=|\mathbf{x}-\mathbf{x}'|$. These metrics are certainly not equivalent, as $\mathbf{H}^{m-1}_{++}$ is unbounded with respect to δ, but with some thought, one sees that they induce the same topology on $\mathbf{H}^{m-1}_{++}$. One can also see that the compact subsets of $\mathbf{H}_{++}^{m-1}$ in this topology are the closed sets for which all the coordinates of all the elements are bounded away from zero (but in fact, any closed subset of $\mathbf{H}_{++}^{m-1}$ already has this property). Finally, with a little work, one can see that δ and Δ are equivalent on such sets. Since we are concerned with the iterates of $\boldsymbol{B}$, we can confine our analysis to $\boldsymbol{B}(\mathscr{H}_+)$. Since $\boldsymbol{B}$ is positive and continuous with respect to Δ and $\mathscr{H}_+$ is compact with respect to Δ, $\boldsymbol{B}(\mathscr{H}_+)$ is a compact subset of $\mathscr{H}_{++}$.

Corollary 1. *Along with the assumptions of Theorem* 2 *as well as the summation conditions of Section* 2, *suppose that* $\varphi \leqslant 3$. *Then* $\boldsymbol{B}$ *has a unique fixed point* $(\mathbf{x}_0, \mathbf{y}_0) \in \mathscr{H}_{++}$, *and if* $(\mathbf{x}(0), \mathbf{y}(0)) \in \mathscr{H}_+$ *is arbitrary*, *then* $(\mathbf{x}(t), \mathbf{y}(t)) \to (\mathbf{x}_0, \mathbf{y}_0)$ *as* $t \to \infty$ *with respect to both* δ *and* Δ.

Proof. First suppose that $\varphi < 3$. If $\varphi = 1$, then Theorem 2 implies that $\boldsymbol{B}$ takes a single value $(\mathbf{x}_0, \mathbf{y}_0) \in \mathscr{H}_{++}$ for all $(\mathbf{x}(0), \mathbf{y}(0)) \in \mathscr{H}_+$ and the corollary is trivial, so suppose that $1 < \varphi < 3$. We then have $0 < 2\frac{\gamma-1}{\gamma+1} < 1$ or $0 < 2\frac{\eta-1}{\eta+1} < 1$. By Theorem 2, $\boldsymbol{B}$ is a contraction mapping on $\mathscr{H}_{++}$ with respect to δ, so we can apply the contraction mapping theorem to the compact and, therefore, complete subspace $\boldsymbol{B}(\mathscr{H}_+)$.

Now suppose that $\varphi = 3$. By Theorem 2 (now denoting elements of $\mathscr{H}_+$ by single vectors for convenience), $\delta(\boldsymbol{B}\mathbf{z}, \boldsymbol{B}\mathbf{z}') < \delta(\mathbf{z}, \mathbf{z}')$ whenever $\mathbf{z}, \mathbf{z}' \in \mathscr{H}_{++}$ are distinct, i.e., $\boldsymbol{B}$ is a weak contraction with respect to δ, and it is clearly continuous on $\mathscr{H}_+$. Also, because δ is a pseudometric, as a map $\mathscr{H}_{++} \to \mathbf{R}$ it is continuous in both variables. Much of the following argument is adapted from [5]. Let $\mathbf{z} \in \mathscr{H}_+$ be arbitrary. Since the sequence $\{\boldsymbol{B}^t\mathbf{z}\}$, $t \geqslant 1$, lies in $\boldsymbol{B}(\mathscr{H}_+)$, it has a convergent subsequence $\{\boldsymbol{B}^{t_l}\mathbf{z}\}$, say $\boldsymbol{B}^{t_l}\mathbf{z} \to \mathbf{v} \in \boldsymbol{B}(\mathscr{H}_+)$. The sequence $\{\delta(\boldsymbol{B}^t\mathbf{z}, \boldsymbol{B}^{t+1}\mathbf{z})\}$ is nonincreasing by weak contractility, so converges to its greatest lower bound and every subsequence has the same limit. Therefore,

$$\begin{aligned}\delta(\mathbf{v}, \boldsymbol{B}\mathbf{v}) &= \delta(\lim \boldsymbol{B}^{t_l}\mathbf{z}, \boldsymbol{B}(\lim \boldsymbol{B}^{t_l}\mathbf{z})) = \delta(\lim \boldsymbol{B}^{t_l}\mathbf{z}, \lim \boldsymbol{B}^{t_l+1}\mathbf{z})\\ &= \lim \delta(\boldsymbol{B}^{t_l}\mathbf{z}, \boldsymbol{B}^{t_l+1}\mathbf{z}) = \lim \delta(\boldsymbol{B}^{t_l+1}\mathbf{z}, \boldsymbol{B}^{t_l+2}\mathbf{z})\\ &= \delta(\boldsymbol{B}\mathbf{v}, \boldsymbol{B}^2\mathbf{v})) = \delta(\boldsymbol{B}\mathbf{v}, \boldsymbol{B}(\boldsymbol{B}\mathbf{v})).\end{aligned}$$

Then, by weak contractility, $\mathbf{v} = \boldsymbol{B}\mathbf{v}$, i.e., $\mathbf{v}$ is a fixed point of $\boldsymbol{B}$. Therefore, $\{\delta(\mathbf{v}, \boldsymbol{B}^t\mathbf{z})\}$ is nonincreasing and converges to its greatest lower bound, which must be 0 since $\delta(\mathbf{v}, \boldsymbol{B}^{t_l}\mathbf{z}) \to 0$, which means that $\boldsymbol{B}^t\mathbf{z} \to \mathbf{v}$. Finally, again by weak contractility, $\mathbf{v}$ is the unique fixed point of $\boldsymbol{B}$. □

As observed in Section 4, the pseudometric δ on $\mathbf{R}_{++}^m$ induces a metric, again called δ, on $\mathbf{P}_{++}^{m-1}$ and, as in that section, we can generalize δ to a metric on $\mathscr{P}_{++} = \mathbf{P}_{++}^{m-1} \times \mathbf{P}_{++}^{n-1}$. There is also a counterpart of Δ on $\mathscr{P}_+ = \mathbf{P}_+^{m-1} \times \mathbf{P}_+^{n-1}$, which may be viewed as depending on the bijection between $\mathscr{H}_+$ and $\mathscr{P}_+$ given by the restriction to $\mathscr{H}_+$ of the natural map $\mathscr{R}_+ \to \mathscr{P}_+$. On $\mathscr{P}_+$, Δ is defined by declaring this bijection an isometry with regard to the two Δ's. This definition of Δ provides one of the usual equivalent metric space structures to $\mathscr{P}_+$. Of course, the further restriction $\mathscr{H}_{++} \to \mathscr{P}_{++}$ is already an isometry with regard to the two δ's. Finally, because it is homogeneous in both coordinates, the bilinear map $\boldsymbol{B}$ also induces a map on $\mathscr{P}_{++}$, which is again called $\boldsymbol{B}$.

Corollary 2. *Along with the assumptions of Theorem* 2, *suppose that* $\varphi \leqslant 3$. *Then* $\boldsymbol{B} : \mathscr{P}_+ \to \mathscr{P}_+$ *has a unique fixed point represented by* $(\mathbf{x}_0, \mathbf{y}_0) \in \mathscr{H}_{++}$, *and if* $(\mathbf{x}(0), \mathbf{y}(0)) \in \mathscr{P}_+$ *is arbitrary*, *then* $(\mathbf{x}(t), \mathbf{y}(t)) = \boldsymbol{B}^t(\mathbf{x}(0), \mathbf{y}(0)) \to (\mathbf{x}_0, \mathbf{y}_0)$ *as* $t \to \infty$ *with respect to both* δ *and* Δ.

Proof. We can apply Corollary 1, identifying $\mathscr{P}_+$ with $\mathscr{H}_+$ via the isometry discussed above, since the summation conditions required in the proof of Corollary 1 were only used to ensure that $\boldsymbol{B} : \mathscr{H}_+ \to \mathscr{H}_+$. □

6. Survival, fecundity, and mating preference

The basic model which was introduced in Section 2 and treated in the last section is only the reproductive portion of a more complete life history model. In this section, we incorporate some other commonly considered life history parameters, requiring us to look at a composite map on the bigger space, $\mathscr{R}_+$.

As in the basic model, each successive generation completely replaces the last. We assume that each female can be and is fertilized once and that half the offspring of each mating is female and half is male. One could allow other proportions of offspring gender with a little more work. Let $\mathbf{x}(t) = (x_1(t), \ldots, x_m(t))$, respectively $\mathbf{y}(t) = (y_1(t), \ldots, y_n(t))$, be the vector of *numbers* of female, respectively male, types in year t. Let $\mathbf{f} \in \mathbf{R}^m_{++}$ be the female fecundity vector, i.e., f_i is the number of eggs generated by a type i female. Let $\mathbf{p} \in \mathbf{H}^{n-1}_+$ be the vector of preferences for males, i.e., p_j is the normalized relative "preference coefficient" exhibited by all females for type j males. Finally, let $\mathbf{r} \in \mathbf{R}^m_{++}$ and $\mathbf{s} \in \mathbf{R}^n_{++}$ be survival vectors for females and males respectively, i.e., r_i and s_j are the proportion of fertilized female eggs of type i and fertilized male eggs of type j respectively which survive to reproduce. Let $\boldsymbol{B}$ be as in Section 2.

The map $\boldsymbol{F}$ on $\mathscr{R}_+$ which transforms one generation into the next is now a composition of maps which is only linear in $\mathbf{x}$. For arbitrary $\mathbf{c} \in \mathbf{R}^m_{++}$, let the function $\mathbf{c} : \mathbf{R}^m_{++} \to \mathbf{R}^m_{++}$ be multiplicative translation by the vector $\mathbf{c}$. Clearly $\mathbf{c}$ is a bijection and for $\mathbf{x}, \mathbf{x}' \in \mathbf{R}^m_{++}$, the definition of d in Section 4 implies that $d(\mathbf{x}, \mathbf{x}') = d(\mathbf{cx}, \mathbf{cx}')$, so $\delta(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{cx}, \mathbf{cx}')$, and $\mathbf{c}$ is an isometry with respect to δ. Just as clearly, $\mathbf{c}$ is linear, in fact diagonal. Let $\pi : \mathbf{R}^n_+ \to \mathbf{H}^{n-1}_+$ be the projection defined by $\pi\mathbf{v} = \mathbf{v}/|\mathbf{v}|$. Then

$$\boldsymbol{F} = (\mathbf{r} \times \mathbf{s}) \circ \tfrac{1}{2}\boldsymbol{B} \circ (\mathbf{f} \times \pi \circ \mathbf{p}).$$

Each of the maps in this composition commutes with the natural map from $\mathscr{R}_+$ to $\mathscr{P}_+$, so $\boldsymbol{F}$ does as well. By Corollary 2, the middle mapping $\frac{1}{2}\boldsymbol{B}$ induces a contraction mapping on $\mathscr{P}_{++}$ with respect to δ if $\varphi \leqslant 3$. Since $\mathbf{f}$, $\mathbf{p}$, $\mathbf{r}$, and $\mathbf{s}$ induce isometries on positive projective space and π induces the identity, $\boldsymbol{F}$ is also a contraction mapping on $\mathscr{P}_{++}$ with the same contraction coefficient as $\boldsymbol{B}$.

We shall find it useful to place a simple and intuitive ordering on $\mathbf{R}^m_+$. If $\mathbf{u} = (u_1, \ldots, u_m)$, $\mathbf{x} = (x_1, \ldots, x_m) \in \mathbf{R}^m_+$, we say $\mathbf{u} \leqslant \mathbf{x}$ if $u_i \leqslant x_i$ and $\mathbf{u} < \mathbf{x}$ if $u_i < x_i$ for $i = 1, \ldots, m$.

Theorem 3. *Let $\boldsymbol{F} : \mathscr{R}_+ \to \mathscr{R}_+$ be defined as above and suppose that $\varphi \leqslant 3$ for the $\boldsymbol{B}$ component of $\boldsymbol{F}$. Then there is a unique $(\mathbf{x}_0, \mathbf{y}_0) \in \mathscr{H}_{++}$ with the property that there are $\kappa, \lambda > 0$ such that $\boldsymbol{F}(\mathbf{x}_0, \mathbf{y}_0) = (\kappa\mathbf{x}_0, \lambda\mathbf{y}_0)$. If $(\mathbf{x}(0), \mathbf{y}(0)) \in \mathscr{R}_+$ is arbitrary and $(\mathbf{x}(t), \mathbf{y}(t)) = \boldsymbol{F}^t(\mathbf{x}(0), \mathbf{y}(0))$ for $t \geqslant 0$, then $(\mathbf{x}(t), \mathbf{y}(t)) \to (\mathbf{x}_0, \mathbf{y}_0)$ in $\mathscr{P}_{++}$ as $t \to \infty$, i.e., $\mathbf{x}(t) \to \mathbf{x}_0$ and $\mathbf{y}(t) \to \mathbf{y}_0$ in direction. Convergence with respect to Δ in $\mathscr{R}_+$ is as follows. If $\kappa > 1$, then $(\mathbf{x}(t), \mathbf{y}(t))$ eventually increases exponentially, i.e., $x_i(t)$ and $y_i(t)$ increase without bound for all i. If $\kappa < 1$, then $(\mathbf{x}(t), \mathbf{y}(t)) \to \mathbf{0}$ eventually exponentially. If $\kappa = 1$ and $\varphi < 3$, then there is a $\mu = \mu(\mathbf{x}(0), \mathbf{y}(0)) > 0$, such that $(\mathbf{x}(t), \mathbf{y}(t)) \to \mu \cdot (\mathbf{x}_0, \lambda\mathbf{y}_0)$, a fixed point of $\boldsymbol{F}$ in $\mathscr{R}_{++}$, implying that λ is the male-to-female ratio at equilibrium in this case. In addition, $\mu : \mathscr{R}_+ \to \mathbf{R}_+$ is linear in $\mathbf{x}$, so is also monotone in $\mathbf{x}$, and is homogeneous of degree zero in $\mathbf{y}$.*

Proof. As noted above, since $\varphi \leqslant 3$, $\boldsymbol{F}$ induces a contraction mapping on $\mathscr{P}_{++}$ with respect to δ. By the contraction mapping theorem, this induced map (technically when restricted to $\boldsymbol{F}(\mathscr{P}_+)$) has a unique fixed point to which all trajectories converge. Translating back to $\mathscr{R}_+$, this fixed point is represented by a unique $(\mathbf{x}_0, \mathbf{y}_0) \in \mathscr{H}_{++}$ and, therefore, there are $\kappa, \lambda > 0$ such that $\boldsymbol{F}(\mathbf{x}_0, \mathbf{y}_0) = (\kappa\mathbf{x}_0, \lambda\mathbf{y}_0)$. Furthermore, if $(\mathbf{x}(0), \mathbf{y}(0)) \in \mathscr{R}_+$ is arbitrary and $(\mathbf{x}(t), \mathbf{y}(t)) = \boldsymbol{F}^t(\mathbf{x}(0), \mathbf{y}(0))$, then $(\mathbf{x}(t), \mathbf{y}(t)) \to (\mathbf{x}_0, \mathbf{y}_0)$ in $\mathscr{P}_{++}$ as $t \to \infty$, in analogy to the portion of the Perron–Frobenius Theorem which follows from Theorem 1. If $\delta((\mathbf{x}(0), \mathbf{y}(0)), (\mathbf{x}_0, \mathbf{y}_0)) = 0$, then $(\mathbf{x}(0), \mathbf{y}(0)) = (a\mathbf{x}_0, b\mathbf{y}_0)$ for some $a, b > 0$ and we can quickly calculate that $(\mathbf{x}(t), \mathbf{y}(t)) = \boldsymbol{F}^t(a\mathbf{x}_0, b\mathbf{y}_0) = a\kappa^{t-1}(\kappa\mathbf{x}_0, \lambda\mathbf{y}_0)$, implying the theorem. Therefore, we may suppose that $\delta((\mathbf{x}(0), \mathbf{y}(0)), (\mathbf{x}_0, \mathbf{y}_0)) > 0$. For $t \geqslant 0$, let $\mu(t), \nu(t)$ be the largest positive numbers such that $\mu(t)\mathbf{x}_0 \leqslant \mathbf{x}(t)$ and $\nu(t)\mathbf{y}_0 \leqslant \mathbf{y}(t)$. Then $\mathbf{x}(t) = \mu(t)(\mathbf{x}_0 + \mathbf{u}(t))$, $\mathbf{y}(t) = \nu(t)(\mathbf{y}_0 + \mathbf{w}(t))$, where $\mathbf{u}(t) \in \mathbf{R}^m_+ \cup \{\mathbf{0}\}$ and $\mathbf{w}(t) \in \mathbf{R}^n_+ \cup \{\mathbf{0}\}$ both have at least one zero coordinate. Let $\mathbf{x}_0 = (x_1, \ldots, x_m)$. Then, recalling the definition of d and noting that $\mathbf{x}(t) \to \mathbf{x}_0$ in δ, we have

$$\begin{aligned} 1 + \max_{1\leqslant i\leqslant m} \frac{u_i(t)}{x_i} &= \max_{1\leqslant i\leqslant m} \left(\frac{(\mathbf{x}_0 + \mathbf{u}(t))_i}{x_i} \right) \\ &= d(\mathbf{x}_0 + \mathbf{u}(t), \mathbf{x}_0) = d(\mathbf{x}(t), \mathbf{x}_0) \to 1. \end{aligned} \tag{2}$$

Therefore, $\frac{u_i(t)}{x_i} \to 0$ for all i, so $|\mathbf{u}(t)| \to 0$ as $t \to \infty$. With an identical argument, we conclude that $|\mathbf{w}(t)| \to 0$ as well. Now suppose that $\kappa > 1$ and choose any κ' such that $1 < \kappa' < \kappa$. Since F_1 is continuous at $(\mathbf{x}_0, \mathbf{y}_0)$ and $F_1(\mathbf{x}_0, \mathbf{y}_0) = \kappa \mathbf{x}_0$, there is an $\varepsilon > 0$ such that whenever $|\mathbf{x} - \mathbf{x}_0|, |\mathbf{y} - \mathbf{y}_0| < \varepsilon$, then $F_1(\mathbf{x}, \mathbf{y}) \geqslant \kappa' \mathbf{x}_0$. Now, let t be so large that $|\mathbf{u}(t)|, |\mathbf{w}(t)| < \varepsilon$. Then, since F_1 is homogeneous of degree one in $\mathbf{x}$ and homogeneous of degree zero in $\mathbf{y}$, we have for such t,

$$\begin{aligned}\mu(t+1)(\mathbf{x}_0 + \mathbf{u}(t+1)) &= \mathbf{x}(t+1) = F_1(\mathbf{x}(t), \mathbf{y}(t))\\ &= F_1(\mu(t)(\mathbf{x}_0 + \mathbf{u}(t)), \nu(t)(\mathbf{y}_0 + \mathbf{w}(t)))\\ &= \mu(t) F_1(\mathbf{x}_0 + \mathbf{u}(t), \mathbf{y}_0 + \mathbf{w}(t))(t) \geqslant \kappa' \mu(t) \mathbf{x}_0\end{aligned}$$

which implies that $\mu(t+1) \geqslant \kappa' \mu(t)$ since $\mathbf{u}(t+1)$ has at least one zero coordinate. Therefore, $\mu(t)$ eventually increases exponentially and so does $\mathbf{x}(t)$. Now, being continuous with respect to Δ, each of the n components of F_2 takes a minimum on the compact set $\mathscr{H}_+$, and since F_2 is positive on $\mathscr{H}_+$, there is a $\mathbf{y}_1 > \mathbf{0}$ such that $F_2(\mathbf{x}, \mathbf{y}) \geqslant \mathbf{y}_1$ for all $(\mathbf{x}, \mathbf{y}) \in \mathscr{H}_+$. But since F_2 is also homogeneous of degree 0 in $\mathbf{y}$, this inequality holds on $\mathbf{H}_+^{m-1} \times \mathbf{R}_+^n$, and since F_2 is linear in $\mathbf{x}$ as well,

$$\begin{aligned}\mathbf{y}(t+1) &= F_2(\mathbf{x}(t), \mathbf{y}(t)) = \mu(t) F_2(\mathbf{x}_0 + \mathbf{u}(t), \mathbf{y}_0 + \mathbf{w}(t))\\ &\geqslant \mu(t) F_2(\mathbf{x}_0, \mathbf{y}_0 + \mathbf{w}(t)) \geqslant \mu(t)\mathbf{y}_1\end{aligned}$$

and $\mathbf{y}(t)$ also eventually increases exponentially. An analogous argument works for $\kappa < 1$.

Now suppose that $\kappa = 1$ and $\varphi < 3$, and let $\omega \leqslant \frac{\varphi - 1}{\varphi + 1} < 1$ be the contraction coefficient for $\boldsymbol{F}$. The remainder of the proof is, unfortunately, even more tedious than what lies above without being enlightening. Taking the logarithm of Eq. (2), we have for $i = 1, \ldots, m$,

$$\begin{aligned}\log\left(1 + \frac{u_i(t)}{x_i}\right) &\leqslant \delta(\mathbf{x}(t), \mathbf{x}_0) \leqslant \delta((\mathbf{x}(t), \mathbf{y}(t)), (\mathbf{x}_0, \mathbf{y}_0))\\ &= \delta(\boldsymbol{F}^t(\mathbf{x}(0), \mathbf{y}(0)), \boldsymbol{F}^t(\mathbf{x}_0, \mathbf{y}_0))\\ &\leqslant \delta((\mathbf{x}(0), \mathbf{y}(0)), (\mathbf{x}_0, \mathbf{y}_0))\omega^t.\end{aligned}$$

We borrow a standard notation and say for any *real* valued function f that $f(t) = \mathrm{O}(\omega^t)$ if there is a $c > 0$ such that $f(t) \leqslant c\omega^t$ for all t. (This does not exclude the possibility that f takes negative values of large absolute value but, in fact, only the last application below of this notation concerns f which can take negative values.) The inequality above implies that $\log\left(1 + \frac{u_i(t)}{x_i}\right) = \mathrm{O}(\omega^t)$. Since $\frac{u_i(t)}{x_i} \to 0$ for all i, for sufficiently large t, $\frac{u_i(t)}{x_i} < 1$. Using the first two terms of the Taylor expansion for log for such t, we see that

$$\frac{u_i(t)}{2x_i} \leqslant \frac{u_i(t)}{x_i}\left(1-\frac{u_i(t)}{2x_i}\right) = \frac{u_i(t)}{x_i} - \frac{1}{2}\left(\frac{u_i(t)}{x_i}\right)^2$$
$$< \log\left(1+\frac{u_i(t)}{x_i}\right) = \mathrm{O}(\omega^t)$$

so $u_i(t) = \mathrm{O}(\omega^t)$ and, summing over i, we conclude that $|\mathbf{u}(t)| = \mathrm{O}(\omega^t)$. Similarly, $|\mathbf{w}(t)| = \mathrm{O}(\omega^t)$. Then, since F_1 is linear in $\mathbf{x}$, homogeneous of degree zero in $\mathbf{y}$, and positive, we have for any $t \geqslant 0$,

$$\begin{aligned}\mu(t+1)(\mathbf{x}_0+\mathbf{u}(t+1)) &= \mathbf{x}(t+1) = F_1(\mathbf{x}(t),\mathbf{y}(t))\\ &= F_1(\mu(t)(\mathbf{x}_0+\mathbf{u}(t)), \nu(t)(\mathbf{y}_0+\mathbf{w}(t)))\\ &= \mu(t)(F_1(\mathbf{x}_0,\mathbf{y}_0+\mathbf{w}(t)) + F_1(\mathbf{u}(t),\mathbf{y}_0+\mathbf{w}(t)))\\ &\geqslant \mu(t)F_1(\mathbf{x}_0,\mathbf{y}_0+\mathbf{w}(t)). \end{aligned} \tag{3}$$

Since $\mathbf{x}_0 > \mathbf{0}$ and F_1 is continuous, arguing as earlier, there is a $c_1 > 0$ such that $F_1(\mathbf{x},\mathbf{y}) \leqslant c_1\mathbf{x}_0$ for all $(\mathbf{x},\mathbf{y}) \in \mathbf{H}_+^{m-1} \times \mathbf{R}_+^n$. Therefore,

$$F_1(\mathbf{u}(t),\mathbf{y}_0+\mathbf{w}(t)) = |\mathbf{u}(t)|F_1(\pi\mathbf{u}(t),\mathbf{y}_0+\mathbf{w}(t)) < \mathrm{O}(\omega^t)c_1\mathbf{x}_0 = \mathrm{O}(\omega^t)\mathbf{x}_0.$$

Hence, dividing equation/inequality (3) by $\mu(t)$, we get

$$\begin{aligned}F_1(\mathbf{x}_0,\mathbf{y}_0+\mathbf{w}(t)) &\leqslant \frac{\mu(t+1)}{\mu(t)}(\mathbf{x}_0+\mathbf{u}(t+1))\\ &= F_1(\mathbf{x}_0,\mathbf{y}_0+\mathbf{w}(t)) + \mathrm{O}(\omega^t)\mathbf{x}_0.\end{aligned} \tag{4}$$

Let $G_1 = \mathbf{r} \circ \frac{1}{2}B_1$, so that $F_1 = G_1 \circ (\mathbf{f} \times \pi \circ \mathbf{p})$. We will exploit the fact that G_1 is linear in $\mathbf{y}$ and is positive. Now,

$$\begin{aligned}F_1(\mathbf{x}_0,\mathbf{y}_0+\mathbf{w}(t)) &= G_1(\mathbf{f}\mathbf{x}_0, \pi(\mathbf{p}\mathbf{y}_0+\mathbf{p}\mathbf{w}(t)))\\ &= G_1\left(\mathbf{f}\mathbf{x}_0, \frac{\mathbf{p}\mathbf{y}_0+\mathbf{p}\mathbf{w}(t)}{|\mathbf{p}\mathbf{y}_0+\mathbf{p}\mathbf{w}(t)|}\right).\end{aligned} \tag{5}$$

Furthermore, since $|\mathbf{w}(t)| = \mathrm{O}(\omega^t)$, so is $|\mathbf{p}\mathbf{w}(t)| = \mathrm{O}(\omega^t)$, so let $c_2 > 0$ satisfy $\frac{|\mathbf{p}\mathbf{w}(t)|}{|\mathbf{p}\mathbf{y}_0|} \leqslant c_2\omega^t$, implying that

$$\begin{aligned}|\mathbf{p}\mathbf{y}_0+\mathbf{p}\mathbf{w}(t)| &= |\mathbf{p}\mathbf{y}_0| + |\mathbf{p}\mathbf{w}(t)| = |\mathbf{p}\mathbf{y}_0|\left(1+\frac{|\mathbf{p}\mathbf{w}(t)|}{|\mathbf{p}\mathbf{y}_0|}\right)\\ &\leqslant |\mathbf{p}\mathbf{y}_0|(1+c_2\omega^t).\end{aligned} \tag{6}$$

Since $\kappa = 1$, we have $F_1(\mathbf{x}_0,\mathbf{y}_0) = \mathbf{x}_0$. Therefore, since G_1 is positive and linear in $\mathbf{y}$, by Eq. (5) and inequality (6),

(a) $F_1(\mathbf{x}_0,\mathbf{y}_0+\mathbf{w}(t)) \geqslant G_1\left(\mathbf{f}\mathbf{x}_0, \dfrac{\mathbf{p}\mathbf{y}_0+\mathbf{p}\mathbf{w}(t)}{|\mathbf{p}\mathbf{y}_0|(1+c_2\omega^t)}\right)$

$$\geqslant G_1\left(\mathbf{fx}_0, \frac{\mathbf{py}_0}{|\mathbf{py}_0|(1+c_2\omega^t)}\right)$$

$$= \frac{1}{(1+c_2\omega^t)} F_1(\mathbf{x}_0, \mathbf{y}_0) = \frac{1}{(1+c_2\omega^t)}\mathbf{x}_0,$$

(b) $$F_1(\mathbf{x}_0, \mathbf{y}_0 + \mathbf{w}(t)) \leqslant G_1\left(\mathbf{fx}_0, \frac{\mathbf{py}_0 + \mathbf{pw}(t)}{|\mathbf{py}_0|}\right)$$

$$= G_1(\mathbf{fx}_0, \pi(\mathbf{py}_0)) + G_1\left(\mathbf{fx}_0, \frac{\mathbf{pw}(t)}{|\mathbf{py}_0|}\right)$$

$$= F_1(\mathbf{x}_0, \mathbf{y}_0) + \frac{|\mathbf{pw}(t)|}{|\mathbf{py}_0|} G_1(\mathbf{fx}_0, \pi\,\mathbf{pw}(t))$$

$$= \mathbf{x}_0 + \frac{|\mathbf{pw}(t)|}{|\mathbf{py}_0|} F_1(\mathbf{x}_0, \mathbf{w}(t))$$

$$\leqslant \left(1 + \frac{|\mathbf{pw}(t)|}{|\mathbf{py}_0|} c_1\right)\mathbf{x}_0 \leqslant (1 + c_1 c_2 \omega^t)\mathbf{x}_0.$$

Then, using these two inequalities along with inequality (4), we obtain

$$\frac{1}{1+c_2\omega^t}\mathbf{x}_0 \leqslant \frac{\mu(t+1)}{\mu(t)}(\mathbf{x}_0 + \mathbf{u}(t+1)) \leqslant (1 + c_1c_2\omega^t + \mathrm{O}(\omega^t))\mathbf{x}_0.$$

But, $\mathbf{u}(t+1) = \mathrm{O}(\omega^{t+1})\mathbf{x}_0$, so from the left inequality we get

$$\frac{1}{1+c_2\omega^t}\mathbf{x}_0 \leqslant \frac{\mu(t+1)}{\mu(t)}(1 + \mathrm{O}(\omega^{t+1}))\mathbf{x}_0,$$

$$\mathbf{x}_0 \leqslant \frac{\mu(t+1)}{\mu(t)}(1 + \mathrm{O}(\omega^t))\mathbf{x}_0,$$

and from the right inequality,

$$\frac{\mu(t+1)}{\mu(t)}\mathbf{x}_0 \leqslant (1 + \mathrm{O}(\omega^t))\mathbf{x}_0$$

which imply, not once but m times, that

$$1 \leqslant \frac{\mu(t+1)}{\mu(t)}(1 + \mathrm{O}(\omega^t)), \tag{7a}$$

$$\frac{\mu(t+1)}{\mu(t)} \leqslant (1 + \mathrm{O}(\omega^t)). \tag{7b}$$

Then these two inequalities imply that for some $c_3 > 0$,

$$\frac{1}{1+c_3\omega^t} \leqslant \frac{\mu(t+1)}{\mu(t)} \leqslant 1 + c_3\omega^t.$$

Applying this inequality t times and multiplying all together, we get

$$\prod_{i=0}^{t-1} \frac{1}{1+c_3\omega^i} \leqslant \frac{\mu(t)}{\mu(0)} \leqslant \prod_{i=0}^{t-1}(1+c_3\omega^i).$$

Then, taking logarithms and using the fact that $\log(1+u) < u$ for $u > 0$, we deduce that for $t \geqslant 1$,

$$-c_3 \sum_{i=0}^{t-1} \omega^i < \log\mu(t) - \log\mu(0) < c_3 \sum_{i=0}^{t-1} \omega^i.$$

Since the geometric series converges, $\{\log\mu(t)\}$ is bounded above and below, which implies that $\{\mu(t)\}$ is both bounded and bounded above 0. Then, multiplying inequalities (7) by $\mu(t)$, we obtain $\mu(t) - \mu(t+1) = \mathrm{O}(\omega^t) = \mu(t+1) - \mu(t)$, so $|\mu(t+1) - \mu(t)| = \mathrm{O}(\omega^t)$, which implies that the series $\sum_{t=0}^{\infty}(\mu(t+1) - \mu(t))$ is absolutely convergent and, therefore, convergent. Then the sequence $\{\mu(t)\}$ is convergent and since it is bounded above 0, $\mu(t) \to \mu$ for some $\mu > 0$. Therefore, $\mathbf{x}(t) \to \mu\mathbf{x}_0$ and, since F_2 is homogeneous of degree 0 in $\mathbf{y}$ and continuous,

$$\begin{aligned}\mathbf{y}(t+1) &= F_2(\mathbf{x}(t), \mathbf{y}(t)) = F_2(\mathbf{x}(t), \mathbf{y}_0 + \mathbf{w}(t)) \to F_2(\mu\mathbf{x}_0, \mathbf{y}_0)\\ &= \mu F_2(\mathbf{x}_0, \mathbf{y}_0) = \mu\lambda\mathbf{y}_0.\end{aligned}$$

Finally, since F_1 is linear in $\mathbf{x}$ and homogeneous of degree zero in $\mathbf{y}$, so is μ. □

7. Concluding remarks

Theorem 2 and its corollaries regarding $\boldsymbol{B}$, a vector valued map with two bilinear components, are the main results of this paper. We hope these will have applications extending well beyond the population model studied here. That model has involved compositions of $\boldsymbol{B}$ with multiplicative translations by fixed vectors in either component. Such a translation is the same as multiplication by a diagonal matrix with positive diagonal entries and is an isometry. More generally, if $\mathbf{A}$ is a square matrix of appropriate size, its composition with either component of $\boldsymbol{B}$, preceding or succeeding, is again bilinear. Suppose that $\boldsymbol{B}$ is a contraction with respect to δ. Then, by Theorem 1, if $\mathbf{A}$ is positive, the composition is also a contraction. In fact, if $\mathbf{A}$ is a "row allowable" nonnegative matrix (every row has at least one nonzero entry), then $\mathbf{A}$ preserves positive vectors and, by a continuity argument using Theorem 1, it is nonexpansive with respect to δ, i.e., $\delta(\mathbf{Ax}, \mathbf{Ax}') \leqslant \delta(\mathbf{x}, \mathbf{x}')$ for all $\mathbf{x}, \mathbf{x}'$. Again, the composition is a contraction.

In Section 6, we spent an inordinate amount of time proving a result about something that almost never happens mathematically, namely the case $\kappa = 1$ exactly and $\varphi < 3$. However, conditions on actual populations often keep them at a steady total size which, one might argue, means that the life history parameters might force κ toward 1.

Although $\varphi \leqslant 3$ is sufficient, it seems far from necessary. Convergence of all trajectories to a single fixed direction occurs in simulations for positive $\mathbf{S}_k$ and $\mathbf{T}_k$ with φ as high as 70, and such simulations have not yet produced any counterexamples.

Acknowledgment

The authors wish to thank Professor Dave Hankin of the Fisheries Department of Humboldt State University for posing the more difficult problem, modelled by a higher order difference system, which spawned the problem discussed here and which we have been unable to solve.

References

[1] G. Birkhoff, Extensions of Jentzsch's theorem, Trans. Amer. Math. Soc. 85 (1957) 219–227.
[2] G. Birkhoff, Lattice Theory, third ed., American Mathematical Society Colloquium Publications, vol. 25, American Mathematical Society, Providence, 1967.
[3] E. Kohlberg, J.W. Pratt, The contraction mapping approach to the Perron–Frobenius theory: why Hilbert's metric? Math. Oper. Res. 7 (2) (1980) 198–210.
[4] E. Kohlberg, A. Neyman, Convergence in Hilbert's metric and convergence in direction, J. Math. Anal. Appl. 93 (1) (1983) 104–108.
[5] E. Kohlberg, The Perron–Frobenius theorem without additivity, J. Math. Econom. 10 (2–3) (1982) 299–303.
[6] S. Gaubert, J. Gunawardena, The Perron–Frobenius theorem for homogeneous, monotone functions, Technical Report HPL-BRIMS-2001-12, Basic Research Institute in the Mathematical Science, HP Laboratories Bristol, 2001.
[7] D. Hilbert, Neue Begründung der Bolyai-Lobatschefskyschen Geometrie, Math. Ann. 57 (1903) 137–150.
[8] E. Seneta, Non-negative Matrices and Markov Chains, second ed., Springer-Verlag, New York, 1981.
[9] R. Cavazos-Cadena, An alternative derivation of Birkhoff's formula for the contraction coefficient of a positive matrix, Linear Algebra Appl. 375 (2003) 291–297.
[10] J.E. Carroll, Birkhoff's contraction coefficient, Linear Algebra Appl. 389 (2004) 227–234.