2 → $n$: Some Tools From Linear Algebra For Analyzing Complex Population Dynamics Models

Lee Altenberg

Twitty Room, 4th Floor, Herrin Labs, November 27, 2012
Outline

1. Intro
2. Karlin’s Thms. 5.1 and 5.2
3. Thm. 5.2: Donsker-Varadhan
4. Multi-Hit
5. Karlin’s Thm 5.1
McNamara and Dall (2011) Model

McNamara and Dall (2011) obtain analytical results for a model of dispersal in random environments:

\[
\begin{bmatrix}
  x_1(t+1) \\
  x_2(t+1)
\end{bmatrix} = M(m) \cdot D
\begin{bmatrix}
  x_1(t) \\
  x_2(t)
\end{bmatrix},
\]

where heterogeneous growth rates are:

\[
D = \begin{bmatrix}
  D_1 & 0 \\
  0 & D_2
\end{bmatrix},
\]

dispersal rate is \( m \), and random environments \( P \) with stationary distribution \( \pi = P\pi \) gives:

\[
M(m) = (1-m) \begin{bmatrix}
  1-P_{21} & P_{12} \\
  P_{21} & 1-P_{21}
\end{bmatrix} + m \begin{bmatrix}
  \pi_1 & \pi_1 \\
  1-\pi_1 & 1-\pi_1
\end{bmatrix}.
\]
Two Theorems of Karlin (1982)


**Theorem (5.1)**

\[ r(M_i M_j D) < r(M_j D) \]

**Theorem (5.2)**

\[ \frac{d}{dm} r \left( [(1 - m) I + m P] D \right) < 0 \]

Assumptions:

5.1: Stochastic irreducible \( M_j = D_1 S_j D_2, M_i M_j = M_j M_i \)

Positive definite \( S_j \),

Positive diagonals \( D_1, D_2, D \neq c I \).

5.2: Stochastic irreducible \( P, 0 \leq m \leq 1, D \neq c I \).
Take-home message of both theorems:

- **Greater mixing reduces growth.**

- Greater mixing may drive some alleles to extinction, thus decreasing genetic diversity.
Karlin’s Theorem 5.2 applies to strongly connected dispersal networks of arbitrary complexity and asymmetry:

**KARLIN, 1982**
Terminology

Let \( \mathbf{A} \) be an \( n \times n \) matrix.

\[
[A]_{ij} = A_{ij} \quad \text{is the \((i, j)\) element of \( \mathbf{A} \).}
\]

\( \mathbf{A}^\top \) is the transpose of \( \mathbf{A} \).

\( \mathbf{I} \) is the identity matrix.

\( \mathbf{D} \) is a diagonal matrix. A positive diagonal “\( \mathbf{D} > 0 \)” means \( D_{ii} > 0 \).

\( \mathbf{e} \) is the vector of 1s.

\( \mathbf{A} \geq 0 \) is irreducible if for each \((i, j)\), some \( t \) gives

\[
[A^t]_{i,j} > 0.
\]

(\( \lambda_i(\mathbf{A}) \)) are the eigenvalues of \( \mathbf{A} \).

(\( \rho(\mathbf{A}) \) := \( \max_i |\lambda_i| \) is the spectral radius of \( \mathbf{A} \).

(\( r(\mathbf{A}) \) := \( \max_i \text{Re}(\lambda_i) \) is the spectral bound of \( \mathbf{A} \).}
Perron-Frobenius Theory for nonnegative irreducible $A$:

1. $\rho(A) = \lambda_1(A) \geq |\lambda_i(A)| > 0$ for all $i$: the **Perron root**

2. $A \mathbf{v}(A) = \rho(A) \mathbf{v}(A) > 0$: **right Perron vector**

3. $\mathbf{u}(A)^\top A = \rho(A) \mathbf{u}(A)^\top > 0$: **left Perron vector**

4. Normalized: $\mathbf{e}^\top \mathbf{v} = \mathbf{u}^\top \mathbf{v} = 1$. 
**Terminology**

A is *essentially nonnegative* if all off-diagonal elements are nonnegative, i.e. there is some $D > 0$ such that $A + D \geq 0$.

Perron-Frobenius Theory for essentially-nonnegative irreducible $A$:

1. $r(A) = \lambda_1(A)$: the *Perron root*
2. $A \mathbf{v}(A) = r(A) \mathbf{v}(A) > 0$: *right Perron vector*
3. $\mathbf{u}(A)\mathbf{u}^\top A = r(A) \mathbf{u}(A)\mathbf{u}^\top > 0$: *left Perron vector*
4. Normalized: $\mathbf{e}^\top \mathbf{v} = \mathbf{u}^\top \mathbf{v} = 1$. 
Donsker-Varadhan Variational Formula for the Spectral Radius

- Karlin’s proof of Theorem 5.2 uses the Donsker and Varadhan (1975) variational formula for the spectral radius.

- Friedland’s (1981) matrix version of Donsker-Varadhan:

\[
r(A) = \sup_{p \in \mathcal{P}_n} \inf_{x > 0} \sum_{i=1}^{n} p_i \frac{[Ax]_i}{x_i} \quad \text{or} \quad \\
\log r(A) = \sup_{p \in \mathcal{P}_n} \inf_{x > 0} \sum_{i=1}^{n} p_i \log \frac{[Ax]_i}{x_i}
\]

where \( A \) is nonnegative, \( \mathcal{P}_n \) is the space of probability vectors, \( p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \).
Donsker-Varadhan: \[ r(A) = \text{saddle-point of } \sum_i p_i \frac{[Ax]_i}{x_i} \]
Friedland’s (1981) add-on to Donsker-Varadhan. Let \( \hat{p} \) and \( \hat{x} \) be the vectors at which the sup and inf are attained. So

\[
r(A) = \sup_{p \in \mathcal{P}_n} \inf_{x > 0} \sum_{i=1}^{n} p_i \frac{[Ax]_i}{x_i} = \sum_{i=1}^{n} \hat{p}_i \frac{[A\hat{x}]_i}{\hat{x}_i}.
\]

They are solved by \( \hat{x} = v(A) \) and \( \hat{p} = u(A) \circ v(A) = [u_i(A)v_i(A)]_{i=1}^{n} \).

\[
r(A) = \sum_{i=1}^{n} u_i v_i \frac{[Av]_i}{v_i} = \sum_{i=1}^{n} u_i v_i \frac{r(A)v_i}{v_i} = \sum_{i=1}^{n} u_i v_i r(A)
\]

\[= r(A).\]
The inf relation implies:

\[
    r(A) = \sup_{p \in P_n} \inf_{x > 0} \sum_{i=1}^{n} p_i \frac{[Ax]_i}{x_i} = \sum_{i=1}^{n} \hat{p}_i \frac{[A\hat{x}]_i}{\hat{x}_i}
\]

\[
    < \sum_{i=1}^{n} \hat{p}_i \frac{[Ax]_i}{x_i}.
\]

for any \( x \neq c\hat{x}, \ c > 0 \) because \( \hat{x} \) is a unique (up to scaling) critical point of

\[
    \phi(x; A) := \sum_{i=1}^{n} \hat{p}_i \frac{[Ax]_i}{x_i}
\]
Let’s put Friedland-Donsker-Varadhan to work

Let \( F(m) = mA + D \). Set

\[
\begin{align*}
\mathbf{u}: & \quad \mathbf{u}^\top F(m) = r(F(m))\mathbf{u}^\top \\
\mathbf{v}: & \quad F(m)\mathbf{v} = r(F(m))\mathbf{v}.
\end{align*}
\]

Then for any \( \mathbf{y} \neq c\mathbf{v} \):

\[
\begin{align*}
r(F(m)) &= \sup_{\mathbf{p} \in \mathcal{P}_n} \inf_{\mathbf{x} > 0} \sum_{i=1}^{n} p_i \left[ (mA + D)x \right]_i x_i \\
&= \sum_{i=1}^{n} u_i v_i \left[ (mA + D)v \right]_i v_i \\
&< \sum_{i=1}^{n} u_i v_i \left[ (mA + D)y \right]_i y_i
\end{align*}
\]
\[ r(F(m)) < \sum_{i=1}^{n} u_i v_i \frac{[(mA + D)y]_i}{y_i} \]

\[ = \sum_{i=1}^{n} u_i v_i \frac{[mAy]_i}{y_i} + \sum_{i=1}^{n} u_i v_i \frac{[Dy]_i}{y_i} \]

\[ = m \sum_{i=1}^{n} u_i v_i \frac{[Ay]_i}{y_i} + \sum_{i=1}^{n} u_i v_i D_i \]

\[ = m \sum_{i=1}^{n} u_i v_i \frac{[Ay]_i}{y_i} + \mathbf{u}^\top \mathbf{D} \mathbf{v} \]

To get something useful out of \( \sum_{i=1}^{n} u_i v_i \frac{[Ay]_i}{y_i} \), what is an obvious choice for \( \mathbf{y} \)?
• Use $y = v(A)$.

• As long as $D \neq c I$, then $v(A) \neq v(m A + D)$.

• Then we get Fact #1:

$$r(F(m)) = r(m A + D) < m \sum_{i=1}^{n} u_i v_i \frac{[Av(A)]_i}{v_i(A)} + u^\top D v$$

$$= m \sum_{i=1}^{n} u_i v_i \frac{r(A)v_i(A)}{v_i(A)} + u^\top D v$$

$$= m r(A) \sum_{i=1}^{n} u_i v_i + u^\top D v$$

$$= m r(A) + u^\top D v.$$
We want to know \( \frac{d}{dm} r(F(m)) \).

We use the formula (Caswell:2000):

\[
\frac{d}{dm} r(F(m)) = u(F(m))^\top \frac{dF(m)}{dm} v(F(m))
\]

Here \( F(m) = mA + D \), so

\[
\frac{dF(m)}{dm} = A.
\]

So

\[
\frac{d}{dm} r(F(m)) = \frac{d}{dm} r(mA + D) = u^\top Av.
\]
We can get replace the $A$ in $u^\top Av$ with something more recognizable:

$$F(m) - D = mA + D - D = mA$$

so

$$A = \frac{1}{m}(F(m) - D)$$

Apply Fact #1: $r(F(m)) < m r(A) + u^\top Dv$,

$$\frac{d}{dm} r(F(m)) = u^\top Av = u^\top \frac{1}{m}(F(m) - D)v$$

$$= \frac{1}{m}(u^\top F(m)v - u^\top Dv) = \frac{1}{m}(r(F(m)) - u^\top Dv)$$

$$< \frac{1}{m}(m r(A) + u^\top Dv - u^\top Dv) = \frac{1}{m} m r(A) = r(A).$$
So we get Karlin’s Theorem 5.2 result:

When $D \neq cI$, then

$$\frac{d}{dm} r(mA + D) < r(A).$$
Karlin’s Theorem 5.2

- Karlin examines the particular case where $A = (P - I)D$, where $P$ is a stochastic matrix.

- So

$$F(m) := m A + D = m(P - I)D + D = [(1 - m)I + mP]D.$$ 

- Hence we get the Karlin’s Theorem 5.2:

$$\frac{d}{dm} r([(1 - m)I + mP]D) < r((P - I)D) = 0$$

since $e^\top (P - I)D = (e^\top - e^\top)D = 0.$
Karlin was interested in $F(m) = [(1 - m)I + mP]D$ because it captures variation in dispersal rates $m$.

But it also captures variation in mutation or recombination rates $m$. 
Application of Karlin’s Theorem 5.2 to the Reduction Principle

For $r \left( [(1 - m)I + mP]D \right)$:

- Suppose the mixing rate $m$
  - is not an extrinsic parameter, but
  - is a variable which is *itself controlled by a gene*.

- A gene which decreases $m$ thus increases $r \left( [(1 - m)I + mP]D \right)$ (by Theorem 5.2), giving itself a growth advantage over its competitor alleles.

Hence, Theorem 5.2 means:

- *Heterogeneous growth selects for reduced mixing.*
Altenberg (1984) showed that Karlin’s Theorem 5.2 explains the repeated appearance of the reduction result in the different contexts, and generalizes the result to a whole class of genetic transmission patterns beyond the special cases that had been analyzed.

- e.g. the cultural transmission of traditionalism
Multiple-Hit Transformations

- \([ (1 - m)I + mP ]D\) represents *single-hit* processes. But

- *Multiple hits* are the norm in many physical and biological situations. E.g.:
  - diffusion with independent advection
  - mutation and recombination at multiple loci
  - combined mutation, recombination, dispersal, etc.

- the variation may not scale the mixing process uniformly (e.g. conditional dispersal, directed mutation)

- Then variation is not of the form \(mA + D\) but rather \(mA + B\), where \(B\) is a linear operator.
Kronecker (Tensor) Product:

\[
\begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix} \otimes \begin{bmatrix}
B_{11} & B_{12} \\
B_{21} & B_{22}
\end{bmatrix} = \begin{bmatrix}
A_{11}B_{11} & A_{11}B_{12} & A_{12}B_{11} & A_{12}B_{12} \\
A_{11}B_{21} & A_{11}B_{22} & A_{12}B_{21} & A_{12}B_{22} \\
A_{21}B_{11} & A_{21}B_{12} & A_{22}B_{11} & A_{22}B_{12} \\
A_{21}B_{21} & A_{21}B_{22} & A_{22}B_{21} & A_{22}B_{22}
\end{bmatrix}
\]
Example: independent mutation at two loci, with selection:

\[ F(m_1, m_2) := \left\{ \left[ (1 - m_1)I_1 + m_1P_1 \right] \otimes \left[ (1 - m_2)I_2 + m_2P_2 \right] \right\} D \]
\[ = \left\{ I_1 \otimes \left[ (1 - m_2)I_2 + m_2P_2 \right] \right\} D \]
\[ + m_1 \left\{ [P_1 - I_1] \otimes \left[ (1 - m_2)I_2 + m_2P_2 \right] \right\} D \]
\[ = B + m_1A \]
\[ \neq D + m_1A \]

where \( P_1 \) is \( n_1 \times n_1 \) and \( P_2 \) is \( n_2 \times n_2 \) (mutation matrices),

\[ A := \left\{ [P_1 - I_1] \otimes \left[ (1 - m_2)I_2 + m_2P_1 \right] \right\} D \]
\[ B := \left\{ I_1 \otimes \left[ (1 - m_2)I_2 + m_2P_2 \right] \right\} D. \]

What do we know in general about the form

\[ \frac{d}{dm} r(mA + B)? \]
For \( y \neq c \ v(m \ A + B) \) the inf relation from Donsker-Varadhan gives

\[
\begin{align*}
 r(m \ A + B) &< \sum_{i=1}^{n} u_i v_i \frac{[(m \ A + B)y]_i}{y_i} \\
&= m \sum_{i=1}^{n} u_i v_i \frac{[Ay]_i}{y_i} + \sum_{i=1}^{n} u_i v_i \frac{[By]_i}{y_i}
\end{align*}
\]

Using the same ‘trick’, \( y = v(A) \):

\[
\begin{align*}
 r(m \ A + B) &< m \ r(A) + \sum_{i=1}^{n} u_i v_i \frac{[Bv(A)]_i}{v_i(A)}
\end{align*}
\]
Fact #1:

\[ r(mA + B) - \sum_{i=1}^{n} u_i v_i \frac{[Bv(A)]_i}{v_i(A)} < m \ r(A). \]

Fact #2: Let’s expand \( r(mA + B) \):

\[ r(mA + B) = u^\top (mA + B)v = m \ u^\top Av + u^\top Bv, \]

to yield:

\[ \frac{d}{dm} r(mA + B) = u^\top Av = \frac{1}{m} [r(mA + B) - u^\top Bv]. \]
Combine our two facts as before:

**Fact #1:** \[ m \ r(A) > r(mA + B) - \sum_{i=1}^{n} u_i v_i \frac{[Bv(A)]_i}{v_i(A)} \]

**Fact #2:** \[ m \frac{d}{dm} r(mA + B) = r(mA + B) - u^\top Bv \]

When \( B = D \), the right-hand sides above are equal and we get \( r(A) > \frac{d}{dm} r(mA + B) \).

But with general \( B \),

\[ ? = \sum_{i=1}^{n} u_i v_i \frac{[Bv(A)]_i}{v_i(A)} \neq u^\top Bv. \]

**We’re stuck (at least I’m stuck).** An open question what to do with this term.
Departures from Reduction

- Examples are known where **departures from reduction** occur, i.e.

\[
\frac{d}{dm} r(mA + B) > r(A).
\]

**Principle of ‘Partial Control’ (Altenberg, 1984)**

*When variation has only partial control over the transformations occurring on types under selection, then it may be possible for that part to evolve an increase in rates.*
Partial Control and Induced Directed Variation
(Altenberg, 2012b)

- Undirected variation of a transformation process,
  i.e. equal scaling of all transition probabilities by a rate $m$,

- may act effectively like directed variation toward fitter types

- due to dynamics induced by other transformation processes and selection,

- so that increases in $m$ increase the population growth rate $r$. 
Enter Karlin’s Theorem 5.1

**Theorem (5.1)**

\[ r(M_iM_jD) < r(M_jD) \]

**Assumptions:**

- Jointly Symmetrizable: \( M_j = D_1S_jD_2 \)
- Positive definite, irreducible \( S_j \)
- Positive diagonal matrices \( D_1, D_2, D \)
- \( D \neq c I \) for any \( c \in \mathbb{R} \)
- Commuting: \( M_iM_j = M_jM_i \)
By using symmetrizable matrices as in Karlin’ Theorem 5.1, results for the form $mA + B$ have been obtained for

- Multilocus mutation rates (Altenberg, 2011), and
- Dispersal in random environments (Altenberg, 2012b).
McNamara and Dall (2011) Model

McNamara and Dall (2011) obtain analytical results for a model of dispersal in random environments:

\[
\begin{bmatrix}
    z_1(t+1) \\
    z_2(t+1)
\end{bmatrix}
= \mathbf{MD}
\begin{bmatrix}
    z_1(t) \\
    z_2(t)
\end{bmatrix},
\]

where heterogeneous growth rates are:

\[
\mathbf{D} = \begin{bmatrix}
    D_1 & 0 \\
    0 & D_2
\end{bmatrix},
\]

dispersal rate is \( m \), and random environments \( \mathbf{P} \) with stationary distribution \( \pi = \mathbf{P}\pi \) gives:

\[
\mathbf{M} = (1-m)\begin{bmatrix}
    1-P_{21} & P_{12} \\
    P_{21} & 1-P_{21}
\end{bmatrix} + m\begin{bmatrix}
    \pi_1 & \pi_1 \\
    1-\pi_1 & 1-\pi_1
\end{bmatrix}.
\]
Theorem (Dispersal in Random Environments (Altenberg, 2012b))

Let $P$ and $Q \in \mathbb{R}^{n,n}$ be transition matrices of reversible ergodic Markov chains that commute with each other. Let $D \neq cI$ be a positive diagonal matrix. Let

$$M(m) := P[(1-m)I + mQ], \quad m \in [0, 1].$$

If all eigenvalues of $P$ are positive, then

$$\frac{d}{dm} r(M(m)D) < 0.$$ 

If all eigenvalues of $P$ other than 1 are negative, then

$$\frac{d}{dm} r(M(m)D) > 0.$$
Multilocus Mutation is a Special Case

- Kronecker products of powers of matrices meet the conditions for Theorem 5.1:
  1. $M_i$ and $M_j$ commute
  2. $M_i$ and $M_j$ are part of a family of symmetrizable matrices $\mathcal{F}$
Define the set of square matrices

\[ \mathcal{M} := \{ A_1, A_2, \ldots, A_L \} \]

where each \( A_i \) is an \( n_i \times n_i \) matrix.

Define

\[ M(t) := \bigotimes_{i=1}^{L} A_i^{t_i} = A_1^{t_1} \otimes A_2^{t_2} \otimes \cdots \otimes A_L^{t_L}, \]

where \( \otimes \) is the Kronecker product (i.e. tensor product), \( t_i \in \{0, 1, 2, \ldots\} \), \( t \in \mathbb{N}_0^L \), \( \mathbb{N}_0 := \{0, 1, 2, \ldots\} \).

Now define the family of such products:

\[ \mathcal{F}(\mathcal{M}) = \left\{ \bigotimes_{i=1}^{L} A_i^{t_i} : t_i \in \{0, 1, 2, \ldots\} \right\}. \]
Kronecker Products of Matrix Powers as a Special Case, cont’d

Clearly, any two members of $\mathcal{F}(\mathcal{M})$ commute, because for any $p, q \in \mathbb{N}_0^L$, then

$$M(p) M(q) = M(q) M(p) = \bigotimes_{i=1}^{L} A_i^{p_i+q_i}.$$

An application where forms $\bigotimes_{i=1}^{L} A_i^{t_i}$ arise is multivariate Markov chains, where $L$ random variables are each independent Markov chain, and $A_i$ is the transition matrix of chain $i$.

In particular the Markov process may be the transmission of information in $L$-symbol strings, which includes the genetic transmission of DNA sequences, for each which $L$ may range up to $L \approx O(10^9)$. 

Lee Altenberg | From 2 to $n$: Analyzing Complex Population Dynamics Models | 38/63
Theorem (Altenberg:2011:Mutation)

The transmission matrix for independent mutation at L loci is:

\[ M_\mu := \bigotimes_{\xi=1}^L \left[ (1 - \mu_\xi)I^{(\xi)} + \mu_\xi P^{(\xi)} \right], \]

where

- \( L \geq 2, \mu \in (0, 1/2)^L, 0 < \mu_\xi < 1/2 \)
- each of the matrices \( P^{(\xi)} \) is a \( n_\xi \times n_\xi \) transition matrix for a reversible ergodic Markov chain,
- \( I^{(\xi)} \) the \( n_\xi \times n_\xi \) identity matrix.
Theorem (Continued)

Let $D \neq c I$. Then for every point $\mu \in (0, 1/2)^L$, the spectral radius of

$$M_\mu D = \left\{ \bigotimes_{\xi=1}^L \left[ (1 - \mu_\xi) I^{(\xi)} + \mu_\xi P^{(\xi)} \right] \right\} D$$

is non-increasing in each $\mu_\xi$.

Note: Condition $\mu_\xi < 1/2$ makes all eigenvalues positive.
Theorem (Concluded)

If diagonal entries

\[ D_{i_1 \ldots i_\xi \ldots i_L} \neq D_{i'_1 \ldots i'_\xi \ldots i'_L} \]

differ for at least one pair \( i_\xi, i'_\xi \in \{1, \ldots, n_\xi\} \), for some

\[
\begin{align*}
    i_1 &\in \{1, \ldots, n_1\}, & i_\xi-1 &\in \{1, \ldots, n_{\xi-1}\}, \\
    i_{\xi+1} &\in \{1, \ldots, n_{\xi+1}\}, & i_L &\in \{1, \ldots, n_L\},
\end{align*}
\]

then

\[
\frac{\partial r(M_\mu D)}{\partial \mu_\xi} < 0.
\]
The key to obtaining the above results is that symmetrizable matrices have a very useful canonical form.
Lemma (Canonical Form for Symmetrizable $M(m)$)

Let $P$ and $Q \in \mathbb{R}^{n,n}$ be transition matrices of ergodic reversible Markov chains that commute with each other. Let

$$M(m) := P[(1-m)I + mQ].$$

Then $P$, $Q$, and $M$ can be decomposed as

$$P = D_\pi^{1/2} K \Lambda_P K^T D_\pi^{-1/2},$$
$$Q = D_\pi^{1/2} K \Lambda_Q K^T D_\pi^{-1/2},$$
$$M(m) = D_\pi^{1/2} K \Lambda_P [(1-m)I + m\Lambda_Q] K^T D_\pi^{-1/2},$$

where $D_\pi \equiv \text{diag}[^\pi]$, $P\pi = Q\pi = \pi$, with $e^T \pi = 1$, $K$ is an orthogonal matrix, and $\Lambda_P$ and $\Lambda_Q$ are diagonal matrices of the eigenvalues of $P$ and $Q$, respectively.
Sources for canonical forms of reversible Markov chain transition matrices:

- Keilson:1979 [p. 33]
- Altenberg:2011:Mutation [Lemmas 1 and 2]
More on the Canonical Form

\[ P = D_{\pi}^{1/2} K \Lambda_P K^\top D_{\pi}^{-1/2} \]
\[ Q = D_{\pi}^{1/2} K \Lambda_Q K^\top D_{\pi}^{-1/2} \]

- \( K \) and \( \pi \) are uniquely determined for any family of commuting symmetrizable stochastic matrices.
- Therefore, the only variation possible for the family is in \( \lambda_i, \ i = 2, \ldots, n \), which means
- there are at most \( n-1 \) degrees of freedom of variation in the family.
Strategy for Using the Canonical Form

- Canonical form $\mathbf{M}(m) = \mathbf{D}_\pi^{1/2} \mathbf{K} \Lambda_m \mathbf{K}^\top \mathbf{D}_\pi^{-1/2}$ is used to produce a symmetric matrix similar to $\mathbf{M}(m)\mathbf{D}$.

- This allows use of the Rayleigh-Ritz formula for the spectral radius.

- The spectral radius simplifies to a sum of squared terms, multiplied by the eigenvalues of the stochastic matrices $\mathbf{P}$ and $\mathbf{Q}$. 
Theorem (Sum-of-Squares Solution for the Spectral Radius)

Let $P$ and $Q \in \mathbb{R}^{n,n}$ be transition matrices of ergodic reversible Markov chains that commute with each other, let $\pi$ be their common right Perron vector, and let $\{\lambda_{Pi}\}$ and $\{\lambda_{Qi}\}$ be their eigenvalues. Let $m \in [0, 1]$ and

$$M(m) := P[(1-m)I + mQ].$$

Let $D$ be a positive diagonal matrix. Set $v \equiv v(M(m)D)$, $u \equiv u(M(m)D)$, $y = (v^\top D_\pi^{-1} Dv)^{-1/2} K^\top D_\pi^{-1/2} D v$. Then

$$\rho(M(m)D) = \sum_{i=1}^{n} \lambda_{Pi} [(1-m) + m\lambda_{Qi} y_i^2].$$
Proof of Sum of Squares Solution

For brevity define $\Phi := K\Lambda_P[(1-m)I + m\Lambda_Q]K^\top$, so $M(m) = B\Phi B^{-1}$.

Since $M(m) \geq 0$ and $B \geq 0$, then $\Phi \geq 0$.

Multiplication by $B$, $D^{1/2}$, and their inverses ($D^{1/2}$ and $D^{-1/2}$ exist since $D$ is a positive diagonal) gives the identities:

$$
\rho(M(m)D) = \rho(B\Phi B^{-1}D) = \rho(\Phi B^{-1}DB) = \rho(\Phi D) = \rho(D^{1/2}\Phi D^{1/2}) = \rho(S),
$$

where $S$ is symmetric:

$$
S := D^{1/2}\Phi D^{1/2} = D^{1/2}K\Lambda_P[(1-m)I + m\Lambda_Q]K^\top D^{1/2}.
$$
Since $\Phi \geq 0$ and $D \geq 0$, then $S \geq 0$.

Since $S$ is symmetric, we may apply the Rayleigh-Ritz variational formula for the spectral radius

$$\rho(S) = \max_{x^\top x = 1} x^\top Sx.$$ 

This yields

$$\rho(M(m)D) = \rho(S) = \max_{x^\top x = 1} x^\top Sx = \max_{x^\top x = 1} x^\top D^{1/2}K\Lambda_P[(1-m)I + m\Lambda_Q]K^\top D^{1/2}x.$$
Since $M$ is irreducible and $D$ a positive diagonal matrix, $MD$ and $S \geq 0$ are irreducible.

So by Perron-Frobenius theory there is a unique eigenvector $\hat{x} > 0$ that yields the maximum in (eq:xVariational).

This allows us to write

$$\rho(M(m)D) = \hat{x}^T D^{1/2} K \Lambda_P [(1-m)I + m\Lambda_Q] K^T D^{1/2} \hat{x}.$$
Define

\[ y := K^T D^{1/2} \hat{x}. \]

Substitution gives:

\[
\rho(M(m)D) = \hat{x}^T D^{1/2} K \Lambda_P [(1 - m)I + m \Lambda_Q] K^T D^{1/2} \hat{x} \\
= y^T \Lambda_P [(1 - m)I + m \Lambda_Q] y \\
= \sum_{i=1}^{n} \lambda_{Pi} [(1 - m) + m \lambda_{Qi}] y_i^2.
\]
An interesting fact to note:

**Theorem (Left and Right Perron Vectors)**

For $M(m) := P[(1-m)I + mQ]$, the left and right Perron vectors of $M(m)D$, $u$ and $v$, are related by

$$u = \frac{1}{(v^\top D^{-1}_\pi D v)} D^{-1}_\pi D v.$$
The general relation is

$$\frac{\partial \rho(A)}{\partial m} = u(A)^\top \frac{\partial A}{\partial m} v(A)$$


The derivative of the spectral radius is thus:

$$\frac{d\rho(MD)}{dm} = \hat{x}^\top \frac{d}{dm} \left[ D^{1/2} K \Lambda_P [(1 - m)I + m \Lambda_Q] K^\top D^{1/2} \right] \hat{x}$$

$$= \hat{x}^\top D^{1/2} K \Lambda_P [\Lambda_Q - I] K^\top D^{1/2} \hat{x}.$$
Derivative of the spectral radius, cont’d

Substitution with \( y := \mathbf{K}^\top \mathbf{D}^{1/2}\hat{\mathbf{x}} \) yields the sum-of-squares form for the derivative:

\[
\frac{d\rho(\mathbf{M}(m)\mathbf{D})}{dm} = \mathbf{y}^\top \Lambda_P (\Lambda_Q - \mathbf{I}) \mathbf{y} = \sum_{i=1}^{n} \lambda_{Pi} (\lambda_{Qi} - 1) y_i^2.
\]
We know several things about the terms in

\[
\frac{d\rho(MD)}{dm} = \sum_{i=1}^{n} \lambda_P(i)(\lambda_{Q_i} - 1)y_i^2
\]

- Since \(P\) and \(Q\) are stochastic matrices, their Perron roots are 1, which here are labelled as \(\lambda_{P1} = \lambda_{Q1} = 1\).
- \(\lambda_{Q1} - 1 = 0\). Thus the first term of the sum is zero.
- \(\lambda_{Qi} - 1 < 0\), for \(i \in \{2, \ldots, n\}\), hence \((\lambda_{Qi} - 1)y_i^2 \leq 0\).
\[ \lambda_{Qi} - 1 < 0, \text{ for } i \in \{2, \ldots, n\} \text{ because:} \]

- Since \( P \) and \( Q \) are symmetrizable, \( \lambda_{Pi}, \lambda_{Qi} \in \mathbb{R} \).

- Since \( P \) and \( Q \) are irreducible, by Perron-Frobenius theory, eigenvalue 1 has multiplicity 1, and \( |\lambda_{Qi}| \leq 1 \), which together imply \( \lambda_{Qi} < 1 \) for \( i \in \{2, \ldots, n\} \).
Derivative of the spectral radius, concluded

- $y_i \neq 0$ for at least one $i \in \{2, \ldots, n\}$, whenever $D \neq cI$ for any $c > 0$. I will skip the details here.

- Therefore $(\lambda_{Qi} - 1)y_i^2 < 0$ for at least one $i \in \{2, \ldots, n\}$.

- If $\lambda_{Pi} > 0$ for all $i$, then $\lambda_{Pi}(\lambda_{Qi} - 1) \leq 0 \ \forall i \geq 2$ so

$$
\frac{d}{dm} \rho(\text{MD}) = \sum_{i=2}^{n} \lambda_{Pi}(\lambda_{Qi} - 1)y_i^2 < 0.
$$

- If $\lambda_{Pi} < 0$ for $i = 2, \ldots, n$, then $\lambda_{Pi}(\lambda_{Qi} - 1) \geq 0 \ \forall i \geq 2$ so

$$
\frac{d}{dm} \rho(\text{MD}) = \sum_{i=2}^{n} \lambda_{Pi}(\lambda_{Qi} - 1)y_i^2 > 0.
$$
Karlin’s Theorem 5.2 gives us the Reduction Principle for single-hit processes.

Karlin’s Theorem 5.1 techniques give us the Reduction Principle for multiple-hit processes; and

Allows derivation of conditions for departure from reduction: negative eigenvalues of the mixing matrix.

Where the Donsker-Varadhan formula for the spectral radius runs of of gas, the properties of symmetrizable matrices can allow us to continue analysis.
Other Applications of Symmetrizable Matrices

Who else has used symmetrizable matrices to advantage?

Population dynamics involve combinations of mixing, $M$, and heterogeneous growth, $D$.

Understanding the relationships between asymptotic growth rates and the properties of populations requires we understand how $r(MD)$ depends on variation in $M$, and variation in $D$.

Karlin’s Theorems 5.1 and 5.2, and their descendants, are principle tools to understand how variation in $M$ maps to variation in $r(MD)$.

Much mathematics remains to be done in this area. I hope you will contribute to its development.

Thank you for your attention!
References


Extension of the Reduction Principle to all *resolvent positive operators*:

**Resolvent positive operators** include familiar examples:

- Reaction-Diffusion models
- Nonlocal Diffusion models
- Laplace-Beltrami operator
- Schrödinger operators
- Heat equation
- 2nd order elliptic operators
- Positive integral operators
Theorem

1. \[ \frac{d}{d\alpha} s(\alpha A + V) \leq s(A), \quad \alpha > 0. \]

2. \( s(\alpha A + V) \) is convex in \( \alpha \);

where

- \( s(\cdot) \) is the spectral bound
- \( A \) is a resolvent positive linear operator on \( C(S) \) or \( L^p(S) \)
- \( V \) is an operator of multiplication in \( C(S) \) or \( L^\infty(S) \)

\( \alpha \) = dispersal rate, \( V \) is heterogeneous growth, \( A \) encompasses most dispersal processes, i.e. ‘mixing’