2005

Cesaro Limits of Analytically Perturbed Stochastic Matrices

Jason Murcko
Harvey Mudd College

Cesaro Limits of Analytically Perturbed Stochastic Matrices

Jason Murcko

Henry Krieger, Advisor

Lesley Ward, Reader

May 11, 2005

HARVEY MUDD
COLLEGE
Department of Mathematics

Abstract

Let $P(\varepsilon) = P_0 + A(\varepsilon)$ be a stochasticity-preserving analytic perturbation of a stochastic matrix P_0. We characterize the hybrid Cesaro limit

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon),$$

where $N(\varepsilon) \uparrow \infty$ as $\varepsilon \downarrow 0$, when P_0 has eigenvalues on the unit circle in the complex plane other than 1.

Contents

Acknowledgments

I would like to thank my advisor, Hank Krieger, for his guidance throughout the past year; without his insight and suggestions at key points, I could not have made the progress I did. Thanks also go to my second reader, Lesley Ward, for her time and helpful commentary. The Reed Institute for Decision Science at Claremont McKenna College funded my early work on this project during the summer of 2004; this experience was very valuable for allowing me to become acquainted with the problem prior to the onset of the hectic school year. Finally, the support and encouragement of family and friends have helped me through the sometimes-stressful thesis process. In particular, I would like to thank my mother for her patient counseling, as well as fellow students Steve Avery, Zajj Daugherty, Brad Greer, Herbie Huff, Eric Malm, and Melissa Strait.

Chapter 1

Introduction

The theory of discrete-time Markov chains on finite state spaces has a wide array of applications in modeling everything from credit ratings to population genetics. The long-term behavior of the chain, which is often of primary interest, is reflected by the behavior of the powers P^n of the associated stochastic matrix of transition probabilities P, as $n \to \infty$. If the recurrent classes of P are all primitive, it is well-known that $\lim_{n\to\infty} P^n$ exists, so this limiting matrix provides the desired long-term information about the chain's behavior. (For definitions of some of the terms used here, refer to Chapter 2.)

On the other hand, if even one of P's recurrent classes is cyclic, this limit will not exist. Instead, the powers of P, as the term cyclic suggests, will tend toward some repeating sequence with finite period. In this case, the *average* long-term behavior of the chain can be represented by the Cesaro limit

$$P^* = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P^k,$$

which is guaranteed to exist and is equal to the eigenprojection for the eigenvalue 1 of P. The Cesaro limit, sometimes referred to as the stationary matrix for P, generalizes $\lim_{n\to\infty} P^n$, as the two are evidently equal if the latter exists.

In the above discussion, we have implicitly assumed that the transition probabilities in P are known exactly. This is often impossible when modeling real-world systems, however, where the probabilities are determined approximately based on observations of how the system in question operates. That is, it will typically be the case that $\hat{p}_{ij} = p_{ij} + \varepsilon_{ij}$; the actual transition probability p_{ij} is estimated by $\hat{p}_{ij}$, and ε_{ij} is an error term, a function that depends on the observations made. To simplify the problem we might suppose that the separate error terms are all analytic functions of a single parameter ε taking on small positive values. We then have an analytically perturbed stochastic matrix $P(\varepsilon) = P_0 + A(\varepsilon)$, where $A(\varepsilon) \to 0$ as $\varepsilon \downarrow 0$ and $P(\varepsilon)$ remains stochastic for all sufficiently small positive ε.

An important problem that arises in this type of situation is to determine the long-term behavior of the perturbed Markov chain, as well as whether this long-term behavior approaches that of the unperturbed Markov chain as the unifying parameter $\varepsilon \downarrow 0$. In other words, we may wonder whether $\lim_{\varepsilon\downarrow 0} P^*(\varepsilon) = P_0^*$, or equivalently if the two limits

$$\lim_{\varepsilon\downarrow 0} \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} P^k(\varepsilon) = \lim_{\varepsilon\downarrow 0} P^*(\varepsilon)$$

and

$$\lim_{N\to\infty} \lim_{\varepsilon\downarrow 0} \frac{1}{N} \sum_{k=1}^{N} P^k(\varepsilon) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} P_0^k = P_0^*$$

are identical. The answer, as it turns out, is no: if the perturbation alters the recurrent-transient structure of the matrix, the two limits will not be equal.

The iterated limits above may lead us to wonder what happens when ε and N are combined in some fashion to form a single hybrid limit. The following expression melds the two limit operations in a fairly straightforward way:

$$\lim_{\varepsilon\downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon).$$

Here $N(\varepsilon)$ takes on positive integer values and increases to ∞ as ε decreases to 0. We can think about this hybrid limit as follows: for each estimate of the actual transition probabilities—that is, each value of ε—we are only interested in the average behavior of the Markov chain up to some finite time limit; but as our estimates improve, we become interested in the chain's behavior over longer and longer spans of time. In [4], Filar, Krieger, and Syed characterize this limit in the case that the unperturbed stochastic matrix P_0 has no eigenvalues on the unit circle in the complex plane other than 1, or equivalently if P_0 has no recurrent classes that are cyclic. More specifically, they show that the hybrid limit exists when $N(\varepsilon) \uparrow \infty$ at different rates, and they show that the value of the limit depends only on properties of the eigenvalue 1 of $P(\varepsilon)$.

In the most general case, though, P_0 may have cyclic recurrent classes, or equivalently eigenvalues on the unit circle other than 1. We investigate the existence and value of the above hybrid Cesaro limit for such cases. In Chapter 2, we introduce some concepts and terminology from matrix theory and linear algebra that provide the framework for much of what comes subsequently; we also describe the structure of a general stochastic matrix and the relationship between this structure and the eigenvalues of the matrix; last, we present an older result concerning stochastic matrix eigenvalues and use it to establish a property of these eigenvalues useful for analyzing the hybrid Cesaro limit. Chapter 3 contains an overview of perturbation theory for matrices. Topics include the way perturbations affect eigenvalues and eigenprojections, a type of reduction process for perturbed matrices that yields information about the perturbed eigenvalues, and some of the peculiarities of perturbed stochastic matrices. The main results concerning the hybrid Cesaro limit are collected in Chapter 4. We first describe how the hybrid Cesaro limit can be decomposed for easier analysis, then present both the previous results from [4] as well as the original results we have obtained. Finally, Chapter 5 reviews aspects of the problem that remain open and challenges in grappling with these open questions.

Chapter 2

Stochastic Matrices and Θ_n

Since the limit operation we are concerned with involves stochastic matrices, the eigenvalues of such matrices appear prominently in our later analysis. In this chapter we present several results relating to stochastic matrix eigenvalues. The most important of these is an original result that yields an estimate crucial for proving statements about the hybrid Cesaro limit.

2.1 Matrix Preliminaries

We begin by reviewing some basic matrix terminology that figures into a number of the concepts introduced subsequently. In what follows, $M_n(\mathbb{C})$ denotes the set of all $n \times n$ matrices with complex entries.

Definition 2.1. *Let $T \in M_n(\mathbb{C})$, and suppose that $\lambda \in \mathbb{C}$ is an eigenvalue of T. The* algebraic multiplicity *of λ is the degree of λ as a root of the characteristic equation* $\det(T - xI) = 0$. *The* geometric multiplicity *of λ is the dimension of the eigenspace*

$$W_\lambda = \{\, v \in \mathbb{C}^n \mid Tv = \lambda v \,\} = \{\, v \in \mathbb{C}^n \mid (T - \lambda I)v = 0 \,\}.$$

Recall that the algebraic multiplicity of an eigenvalue is always at least as large as the geometric multiplicity, but may be strictly larger.

Definition 2.2. *An eigenvalue λ of the matrix $T \in M_n(\mathbb{C})$ is* semisimple *if the algebraic and geometric multiplicities of λ are equal.*

Hence semisimplicity means, in a sense, that the eigenspace for λ is not "deficient." Eigenspaces that are deficient, however, can be repaired by relaxing the definition as follows.

Definition 2.3. *If λ is an eigenvalue of $T \in M_n(\mathbb{C})$, the* generalized eigenspace *associated with λ is the set*

$$\{\, v \in \mathbb{C}^n \mid (T - \lambda I)^k v = 0 \text{ for some } k \in \mathbb{Z}^+ \,\}.$$

Evidently the generalized eigenspace associated with λ contains the eigenspace associated with λ. Additionally, the generalized eigenspace is a subspace of $\mathbb{C}^n$, in the case that we are working with $M_n(\mathbb{C})$, and the dimension of this subspace is equal to the algebraic multiplicity of λ. Hence we can speak simply of the multiplicity of an eigenvalue in reference to this common value.

The sense in which generalized eigenspaces "fix" the type of deficiency mentioned above can be formalized using the notion of direct sums.

Definition 2.4. *Let $W_1, W_2, \ldots, W_m$ be subspaces of the vector space V. Then V is the* direct sum *of the W_i, written*

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_m = \bigoplus_{i=1}^{m} W_i,$$

if for each $v \in V$ there exist unique $w_i \in W_i$, $i = 1, 2, \ldots, m$, such that $v = w_1 + w_2 + \cdots + w_m$. In this situation, the component *of $v \in V$ in W_i is the unique $w_i \in W_i$ that appears in the above decomposition of v.*

If $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of a matrix $T \in M_n(\mathbb{C})$ and $W_1, \ldots, W_m$ are the associated generalized eigenspaces, then it is always the case that $\mathbb{C}^n = \bigoplus_{i=1}^m W_i$. This statement does not hold in general, however, if we substitute eigenspaces for generalized eigenspaces. In particular, it fails precisely when any eigenvalue is not semisimple, or equivalently when T is not diagonalizable. In other words, if any eigenvalue of T is not semisimple then the union of bases for the individual eigenspaces does not span all of $\mathbb{C}^n$.

The discussion of direct sums motivates the following.

Definition 2.5. *Let $T \in M_n(\mathbb{C})$. Then T is a* projection matrix *if $T^2 = T$.*

The notion of a projection can be straightforwardly generalized to linear transformations on arbitrary vector spaces, but for our purposes the definition given is sufficient.

It is a fact that if $\mathbb{C}^n = \bigoplus_{i=1}^m W_i$ and $1 \le i \le m$, there is a unique $P_i^* \in M_n(\mathbb{C})$ satisfying the following two properties:

(1) P_i^* maps $\mathbb{C}^n$ onto W_i; that is, the linear transformation on $\mathbb{C}^n$ defined by $v \mapsto P_i^* v$ has range W_i;

(2) for each $v \in \mathbb{C}^n$, the component of v in W_i is equal to $P_i^* v$.

This P_i^* is a projection matrix: if the decomposition of $v \in \mathbb{C}^n$ based on the W_i is $v = w_1 + \cdots + w_m$, then the component of w_i in W_j is clearly w_i if $i = j$ and 0 if $i \neq j$; thus

$$(P_i^*)^2 v = P_i^* w_i = w_i = P_i^* v,$$

so P_i^* and $(P_i^*)^2$ act identically on $\mathbb{C}^n$ and must be the same matrix. P_i^* is called the projection matrix onto W_i along $W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_m$. In this situation it is straightforward to see that $\sum_{i=1}^m P_i^* = I$. Additionally, $P_i^* P_j^* = \delta_{ij} P_i^*$, where δ_{ij} is the Kronecker delta defined by

$$\delta_{ij} = \begin{cases} 0, & \text{if } i \neq j; \\ 1, & \text{if } i = j. \end{cases}$$

Conversely, if T is a projection matrix, then there are unique subspaces W_1 and W_2 of $\mathbb{C}^n$ such that T is the projection matrix onto W_1 along W_2.

Another useful concept relating to projection matrices is the following.

Definition 2.6. *Suppose that T is the projection matrix onto W_1 along W_2 and that T' is the projection matrix onto W_1' along W_2'. Then T' is a* sub-projection *of T if there is a subspace V of $\mathbb{C}^n$ such that $W_1 = W_1' \oplus V$ and $W_2' = W_2 \oplus V$.*

Using the same type of direct sum decomposition techniques used earlier, it is not difficult to show that if T' is a sub-projection of T, then $TT' = T' = T'T$.

Placing the above discussion of projection matrices in the context of eigenvalues and eigenspaces yields another important piece of terminology.

Definition 2.7. *Let $T \in M_n(\mathbb{C})$, and suppose that $\lambda_1, \ldots, \lambda_m$ and $W_1, \ldots, W_m$ are the eigenvalues and respective generalized eigenspaces for T. The* eigenprojection *for T associated with λ_i, written $P^*(\lambda_i)$, is the projection matrix onto W_i along $W_1, \ldots, W_{i-1}, W_{i+1}, \ldots, W_m$.*

Observe that $TP^*(\lambda_i) = P^*(\lambda_i)T$ for each i: since T maps each W_i to itself, if $v = w_1 + \cdots + w_m$ is the decomposition of v based on the W_i then

$$P^*(\lambda_i)Tv = P^*(\lambda_i)(Tw_1 + \cdots + Tw_m) = Tw_i = TP^*(\lambda_i)v.$$

Thus $P^*(\lambda_i)T$ and $TP^*(\lambda_i)$ act identically on $\mathbb{C}^n$ and are the same matrix.

Finally, recall that a matrix M is *nilpotent* if $M^k = 0$ for some $k > 0$; the smallest positive integer k for which $M^k = 0$ is called the *index* of the nilpotent. In the situation described in the previous paragraph, there is a unique nilpotent matrix $D(\lambda_i)$ which satisfies $TP^*(\lambda_i) = \lambda_i P^*(\lambda_i) + D(\lambda_i)$ and $P^*(\lambda_i)D(\lambda_i) = D(\lambda_i) = D(\lambda_i)P^*(\lambda_i)$. The index of $D(\lambda_i)$ is greater than 1 precisely when λ_i is not semisimple; equivalently, $D(\lambda_i) = 0$ if and only if λ_i is semisimple.

2.2 Stochastic Matrices and Their Structure

Recall that a square matrix is stochastic if its entries are real and nonnegative and the sum of the entries in each row is equal to 1. In this section we relate the structure of a stochastic matrix—that is, the location of its positive entries—to information about its eigenvalues. Most of this material is drawn from Chapter 1 of [9]. Throughout, $P = [p_{ij}]$ denotes a fixed $n \times n$ stochastic matrix and $p_{ij}^{(m)}$ is the ijth entry of P^m.

Definition 2.8. *If i and j are indices from the set $\{1, 2, \dots n\}$, then i* has access *to j, written $i \to j$, if $p_{ij}^{(m)} > 0$ for some $m > 0$. Also, i* communicates with *j, written $i \leftrightarrow j$, if i has access to j and j has access to i.*

Since every row of a stochastic matrix has a positive entry, it is clear that every index has access to some index. On the other hand, it is possible for an index not to communicate with any index. For example, if

$$P = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix},$$

then the index 2 does not communicate with any index since $P^m = P$ for all $m > 0$.

Viewing $\to$ and $\leftrightarrow$ as binary relations on the set of indices, the former is transitive: if $p_{ij}^{(m)} > 0$ and $p_{jk}^{(m')} > 0$, then since $P^{m+m'} = P^m P^{m'}$ we have

$$p_{ik}^{(m+m')} = \sum_{l=1}^{n} p_{il}^{(m)} p_{lk}^{(m')} \geq p_{ij}^{(m)} p_{jk}^{(m')} > 0.$$

The second binary relation has even nicer properties if we restrict our attention to the set of indices $J_c = \{\, i \in \{1, \dots, n\} \mid i \leftrightarrow j \text{ for some index } j \,\}$. Specifically, $\leftrightarrow$ is an equivalence relation on J_c. (That $\leftrightarrow$ is symmetric is evident; transitivity follows from the transitivity of $\to$, and we obtain reflexivity from symmetry and transitivity together with the definition of J_c.) We may thus partition J_c into "communicating classes" of indices, call them $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_k$. Typically, we also group all non-communicating indices into a class of their own, denoted $\mathcal{C}_0$.

We can further classify the set of indices based on the following definition.

Definition 2.9. *An index i is* recurrent *if, for every index j to which i has access, j also has access to i. Otherwise, i is* transient.

Based on what was noted previously, every index in $\mathcal{C}_0$ is transient. For any other $\mathcal{C}_i$, the properties of $\to$ and $\leftrightarrow$ imply that if one index in $\mathcal{C}_i$ is recurrent, all other indices in $\mathcal{C}_i$ are also recurrent. It follows that either every index in $\mathcal{C}_i$ is recurrent or every index in $\mathcal{C}_i$ is transient. Therefore it makes sense to refer to the classes themselves as recurrent or transient based on the classification of their indices. It is a fact that any stochastic matrix contains at least one recurrent class of indices (see [9], p. 16). We now illustrate these concepts with an example.

Example 2.1. Let

$$P = \begin{bmatrix} 0 & 0.3 & 0 & 0.1 & 0.5 & 0.1 & 0 \\ 0 & 0.4 & 0 & 0.1 & 0.2 & 0 & 0.3 \\ 0.2 & 0.7 & 0 & 0 & 0 & 0 & 0.1 \\ 0 & 0.2 & 0 & 0.5 & 0 & 0.2 & 0.1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

One can verify that $\mathcal{C}_0 = \{1, 3\}$ is the class of non-communicating indices, and that among the communicating indices the classes are $\mathcal{C}_1 = \{2, 4\}$, $\mathcal{C}_2 = \{5\}$, and $\mathcal{C}_3 = \{6, 7\}$. Notice that even though the index 1 is non-communicating, $3 \to 1$. Of the communicating classes, $\mathcal{C}_1$ is transient since the index 2, for example, has access to the indices 5 and 7, but not vice versa. On the other hand, both $\mathcal{C}_2$ and $\mathcal{C}_3$ are recurrent: the indices in each class do not have access to any index outside of their own class.

The relationships between indices and classes of indices can be clarified, within a given stochastic matrix, by permuting the indices so that the indices in each class are adjacent. We also typically order the classes with recurrent (communicating) classes first, followed by transient communicating classes, and finally the non-communicating class (if there is one), since the resulting matrix is then lower block triangular—that is, all the positive entries occur either within or below the diagonal blocks associated with the different classes. If we perform this process on the matrix from Example 2.1, placing the classes in the order $\mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_1, \mathcal{C}_0$, we obtain

$$P' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0.3 & 0.4 & 0.1 & 0 & 0 \\ 0 & 0.2 & 0.1 & 0.2 & 0.5 & 0 & 0 \\ 0.5 & 0.1 & 0 & 0.1 & 0.3 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & 0.7 & 0.2 & 0 \end{bmatrix}.$$

Note that this type of reordering is not unique: we can change the order of the indices within each class, as well as the order of the recurrent classes and the order of the transient communicating classes, and the resulting matrix will still be lower block triangular. It is important, however, that whichever particular way the indices are permuted, the underlying structure of the matrix is not affected. To be precise, the original matrix and the permuted one are always similar, so their eigenvalues are the same. In the example with which we have been working, we have the identity $P' = SPS^{-1}$, where

$$S = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Observe that S is a *permutation matrix*: each row and column has exactly one entry equal to 1 and the rest equal to 0. It is not difficult to see that however the indices of P are permuted, the resulting matrix will be similar to P via conjugation by a permutation matrix.

Aside from helping to bring out the inherent structure of a stochastic matrix, the index classification we have described is useful because it yields eigenvalue information about P. For one, from rearranging the indices as above we can see that the eigenvalues of P will be the union of the eigenvalues for each diagonal block. As it turns out, the blocks for the transient classes, as well as the non-communicating class, all have eigenvalues with modulus strictly less than 1 (this follows from Theorem 1.5 in [9], p. 22). On the other hand, for each recurrent class there is a positive integer k such that the associated diagonal block has all the kth roots of unity as eigenvalues of multiplicity 1, and no other eigenvalues on or outside the unit circle in the complex plane (this follows from Theorem 1.1, Corollary 1, and Theorem 1.7 in [9], pp. 3–4, 8, and 23, respectively). If $k = 1$ the class is called *primitive*, whereas if $k > 1$ the class is called *cyclic*. These results imply that if P is a stochastic matrix and λ is an eigenvalue of P satisfying $|\lambda| = 1$, then λ is semisimple. Additionally, 1 is an eigenvalue of P.

2.3 Stochastic Matrix Eigenvalues and Θ_n

Additional information about the eigenvalues of stochastic matrices is available from more sophisticated results.

Definition 2.10. *Let $n \in \mathbb{N}$. We denote by Θ_n the set of all eigenvalues of $n \times n$ stochastic matrices. That is,*

$$\Theta_n = \{\, \lambda \in \mathbb{C} \mid \textit{there is an } n \times n \textit{ stochastic matrix with } \lambda \textit{ as an eigenvalue} \,\}.$$

Trivially, then, $\Theta_1 = \{1\}$ (there is only one 1×1 stochastic matrix), and it is not difficult to show that $\Theta_2 = [-1, 1]$. The following three results can all be found in [7]; the first is due to Dmitriev and Dynkin, the second to Karpelevič. For $x \in \mathbb{R}$, $[x]$ denotes the greatest integer less than or equal to x.

Theorem 2.1. *(*[7], p. 175*) If z is a complex number whose argument is between* 0 *and $2\pi/n$, inclusive, where $n \geq 3$, then z is an eigenvalue of some $n \times n$ stochastic matrix if and only if z lies in the triangle with vertices* 0, 1, *and $e^{2\pi i/n}$.*

Theorem 2.2. *(*[7], pp. 176-177*) Let $n \in \mathbb{N}$. Then Θ_n is symmetric with respect to the real line (that is, $z \in \Theta_n$ if and only if $\overline{z} \in \Theta_n$), is contained within the closed unit disk, and intersects the unit disk's boundary, $|z| = 1$, at precisely the kth roots of unity for $k \leq n$. The boundary of Θ_n consists of these points and of curvilinear arcs connecting them in circular order. For $n > 3$, these arcs can be characterized as follows.*

Consider a boundary arc of Θ_n with endpoints λ' and λ'' in counterclockwise order. Write λ' as $e^{2\pi i(a'/b')}$, where $0 \leq a' < b' \leq n$ and $\gcd(a', b') = 1$. (In particular, when $\lambda' = 1$ this gives $a' = 0$ and $b' = 1$.) Similarly write λ'' as $e^{2\pi i(a''/b'')}$. Then either

$$b'' \left[\frac{n}{b''}\right] \geq b' \left[\frac{n}{b'}\right] \tag{2.1}$$

or

$$b'' \left[\frac{n}{b''}\right] \leq b' \left[\frac{n}{b'}\right], \tag{2.2}$$

and if (2.1) holds for a given arc then (2.2) must hold for the complex conjugate arc (and vice versa). Hence, by the symmetry of Θ_n, we may suppose that (2.1) is satisfied and that $b'' > 1$. (If $b'' = 1$, then $a'' = 0$ and $a'/b' = (n-1)/n$, so that both (2.1) and (2.2) hold for the arc; hence both hold for the complex complex conjugate arc, and we may consider it instead.)

Now let $r_1 = b''$, $r_2 = a''$, and let $r_3, \ldots, r_m$ be the remainders obtained by iteratively using the Euclidean algorithm: $r_k = q_k r_{k+1} + r_{k+2}$, with $0 < r_{k+2} < r_{k+1}$, for $k = 1, 2, \ldots, m-2$, and $r_{m-1} = q_{m-1} r_m$. If $[n/b''] = 1$ and m is even, then the arc is implicitly parametrized by the equation

$$z^q (z^p - t)^r = (1-t)^r, \tag{2.3}$$

where the real parameter t runs over the interval $[0, 1]$, $r = r_{m-1}$, and p and q are defined by

$$a'' p \equiv 1 \pmod{b''}, \qquad 0 < p < b'' \tag{2.4}$$

$$a'' q \equiv -r \pmod{b''}, \qquad 0 \leq q < b''. \tag{2.5}$$

Otherwise, the arc is analogously parametrized by the equation

$$(z^b - t)^d = (1-t)^d z^q, \tag{2.6}$$

where $d = [n/b'']$, $b = b''$, and q is defined by

$$a'' q \equiv -1 \pmod{b''}, \qquad 0 < q < b''. \tag{2.7}$$

Corollary 2.1. *(*[7], p. 175*) The set Θ_3 consists of the points on the interior or boundary of the triangle with vertices 1, $e^{2\pi i/3}$, and $e^{-2\pi i/3}$ together with the points in the interval $[-1, 1]$.*

Using Theorem 2.2, I have been able to prove that Θ_n possesses geometric properties useful in analyzing Cesaro limits.

Lemma 2.1. *Let $p \in \mathbb{C}[z, t]$, viewed as a function from $\mathbb{C} \times \mathbb{R}$ to $\mathbb{C}$ (that is, p is a polynomial in $z = x + iy$ and t). Suppose that $p(z_0, t_0) = 0$, and let*

$$w_1 = \frac{\partial p}{\partial x}(z_0, t_0) \quad \text{and} \quad w_2 = \frac{\partial p}{\partial t}(z_0, t_0).$$

If $w_1 \neq 0$, then the equation $p(z, t) = 0$ implicitly defines a C^1 function $f : I \to \mathbb{C}$, where I is an open interval containing t_0, and $f'(t_0) = -\overline{w}_1 w_2 / |w_1|^2$.

Proof. If we view p as a function from $\mathbb{R}^2 \times \mathbb{R}$ to $\mathbb{R}^2$, then p is certainly C^1. Letting $w_1 = a + bi$ and $w_2 = c + di$, and noting that

$$\frac{\partial p}{\partial y} = \frac{\partial p}{\partial z}\frac{\partial z}{\partial y} = \frac{\partial p}{\partial z} \cdot i = i\left(\frac{\partial p}{\partial z} \cdot 1\right) = i\left(\frac{\partial p}{\partial z}\frac{\partial z}{\partial x}\right) = i\frac{\partial p}{\partial x},$$

we see that

$$\frac{\partial p}{\partial y}(z_0, t_0) = i\frac{\partial p}{\partial x}(z_0, t_0) = iw_1 = -b + ai.$$

Hence the matrix for the derivative of p at (z_0, t_0) is

$$Dp(z_0, t_0) = \begin{bmatrix} a & -b & c \\ b & a & d \end{bmatrix}.$$

Then by the implicit function theorem, we can assert the existence of a C^1 function f from some open interval containing t_0 to $\mathbb{C}$ as long as the matrix

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

is invertible. But this is the case if and only if its determinant, $a^2 + b^2$, is nonzero, and clearly this is true if and only if $w_1 \neq 0$.

Finally, under this assumption the implicit function theorem gives that

$$\begin{aligned} f'(t_0) &= -\begin{bmatrix} a & -b \\ b & a \end{bmatrix}^{-1} \begin{bmatrix} c \\ d \end{bmatrix} = -\frac{1}{a^2+b^2} \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} \\ &= -\frac{1}{a^2+b^2} \begin{bmatrix} ac+bd \\ ad-bc \end{bmatrix}. \end{aligned}$$

It just remains to note that

$$-\frac{\overline{w}_1 w_2}{|w_1|^2} = -\frac{(a-bi)(c+di)}{|a+bi|^2} = -\frac{(ac+bd)+(ad-bc)i}{a^2+b^2}.$$

□

We now arrive at the two original results which later bear on our analysis of Cesaro limits. Throughout these results, we make use of the notation from Theorem 2.2. Before presenting them, we introduce a piece of notation and a concept from number theory. First, if a and b are integers we write $b \mid a$ if b divides a; that is, $b \mid a$ if there is an integer c such that $a = bc$. If b does not divide a, we likewise write $b \nmid a$. Next, suppose that n is a positive integer. The nth *Farey series* $\mathcal{F}_n$ is the increasing sequence of irreducible fractions between 0 and 1 inclusive with denominators no larger than n. For example, $\mathcal{F}_3$ is the sequence $0/1, 1/3, 1/2, 2/3, 1/1$. A property of Farey series useful to us is the following: if a/b and c/d are consecutive fractions in the Farey series $\mathcal{F}_n$, then $bc - ad = 1$ (see [5], Theorem 28, p. 23).

Theorem 2.3 (Krieger-Murcko). *Fix an $n > 3$. Suppose that $\lambda' = e^{2\pi i(a'/b')}$ and $\lambda'' = e^{2\pi i(a''/b'')}$ are consecutive roots of unity as in Theorem 2.2 and that the boundary arc of Θ_n connecting λ' and λ'' is described by (2.3). Then at the endpoints λ' and λ'' of the boundary arc, the arc has well-defined tangent lines l_1' and l_1'', respectively. Also, if l_2' and l_2'' denote the lines tangent to the unit circle $|z| = 1$ at λ' and λ'', respectively, then l_1' makes a nonzero angle with l_2' at λ', and likewise for l_1'' and l_2'' at λ''.*

Proof. To begin, we examine the case where $\lambda' = 1$ and $\lambda'' = e^{2\pi i/n}$. (It is straightforward to see that these are, in fact, consecutive roots of unity.) Then $a'/b' = 0/1$ and $a''/b'' = 1/n$, so $r_1 = b'' = n$, $r_2 = a'' = 1 = r_m$, and $[n/b''] = [n/n] = 1$. Therefore this case satisfies the requirements from Theorem 2.2 for the boundary arc to be described by (2.3). The desired result here follows immediately from Theorem 2.1 since the arc is simply the straight line segment connecting 1 and $e^{2\pi i/n}$. So we may henceforth assume that $a' > 0$, $b' > 1$, and $a''/b'' \neq 1/n$.

We next show that $a'' > 1$ and $q > 0$ for these remaining cases. Suppose first that $a'' = 1$. Since we are assuming that $a''/b'' \neq 1/n$, it follows that $b'' < n$; moreover, as $[n/b''] = 1$, we have that $n \leq 2b'' - 1$. Now, any rational number c/d, where $2 \leq c < d \leq n$ and c and d are relatively prime, satisfies

$$\frac{c}{d} \geq \frac{2}{n} \geq \frac{2}{2b'' - 1} > \frac{2}{2b''} = \frac{1}{b''}. \tag{2.8}$$

Since a'/b' is the greatest rational number less than $a''/b'' = 1/b''$ with denominator less than or equal to n, (2.8) implies that $a' = 1$, whence also $b' = b'' + 1 \leq n$. Finally, we obtain that

$$b'' \left[\frac{n}{b''}\right] = b'' < b'' + 1 = b' = b' \left[\frac{n}{b'}\right].$$

But this inequality contradicts (2.1), so we may conclude that $a'' > 1$.

Now suppose, again for the sake of contradiction, that $q = 0$. Since a'' and b'' are relatively prime, r_m is always equal to 1; thus $r = r_{m-1}$ is the last $r_i > 1$, and if some $r_i > 1$ then $r \leq r_i$. In particular, $r \leq b''$ as $r_1 = b'' > 1$. But $r \equiv_{b''} -a''q = 0$ by (2.5), so it must be the case that $r = b''$. Thus

$r_{m-1} = r_1$, so $a'' = r_2 = r_m = 1$. This contradicts what was just established about a'', so $q > 0$.

Moving on from these preliminaries, let $f(z,t) = z^q(z^p - t)^r - (1-t)^r$, $z_0 = e^{2\pi i(a''/b'')}$, and $z_1 = e^{2\pi i(a'/b')}$. The congruence relation $a''p \equiv_{b''} 1$ from (2.4), together with the fact that $b'' > 1$, implies that

$$z_0^p = e^{2\pi i(a''p/b'')} = e^{2\pi i/b''} \neq 1. \tag{2.9}$$

Hence $f(z_0, 1) = z_0^q(z_0^p - 1)^r \neq 0$. Therefore it must be the case that $f(z_0, 0) = 0$ and $f(z_1, 1) = 0$. Additionally, observe that

$$\frac{\partial f}{\partial x} = qz^{q-1}(z^p - t)^r + prz^{p+q-1}(z^p - t)^{r-1} \tag{2.10}$$

and

$$\frac{\partial f}{\partial t} = r(1-t)^{r-1} - rz^q(z^p - t)^{r-1}. \tag{2.11}$$

We now consider the endpoint z_0. From (2.10) we see that

$$\frac{\partial f}{\partial x}(z_0, 0) = qz_0^{q-1}(z_0^p)^r + prz_0^{p+q-1}(z_0^p)^{r-1} = (pr+q)z_0^{pr+q-1}.$$

But $z_0^{pr+q} - 1 = f(z_0, 0) = 0$, so in fact

$$\frac{\partial f}{\partial x}(z_0, 0) = (pr+q)z_0^{-1}.$$

As p, q, and r are all positive integers, this shows that $\partial f/\partial x$ is nonzero at $(z_0, 0)$. Next let $w_1 = (pr+q)z_0^{-1} = (pr+q)\overline{z}_0$ and

$$w_2 = \frac{\partial f}{\partial t}(z_0, 0) = r - rz_0^q(z_0^p)^{r-1} = r - rz_0^{pr+q-p} = r(1 - z_0^{-p}).$$

Lemma 2.1 implies that if $v = -w_1\overline{w}_2/|w_1|^2$ is nonzero, then the boundary arc implicitly described by (2.3) has a well-defined tangent line at $z_0 = \lambda''$ and v is a vector (complex number) in the direction of this line. Certainly $v \neq 0$ if the dot product of v with the radial vector for z_0 is nonzero, so we begin by showing this. Now, the dot product of two vectors v_1 and v_2 in $\mathbb{R}^2$, when we view v_1 and v_2 as complex numbers, is easily seen to be equal to $\operatorname{Re}(v_1\overline{v}_2)$. Since

$$v\overline{z}_0 = -\frac{\overline{w}_1 w_2}{|w_1|^2}\overline{z}_0 = -\frac{\left[(pr+q)z_0\right]\left[r(1-z_0^{-p})\right]}{(pr+q)^2}\overline{z}_0 = -\frac{r(1-z_0^{-p})}{pr+q},$$

the desired dot product is equal to

$$-\frac{r}{pr+q}(1 - \mathrm{Re}\,(z_0^{-p})).$$

But r and $pr+q$ are both positive, so it only remains to note that z_0^{-p} lies on the unit circle and is not equal to 1 by (2.9). Thus the boundary arc does indeed have a well-defined tangent line l_1'' at z_0.

To establish that l_1'' makes a nonzero angle with l_2'', the line tangent to the unit circle at $z_0 = \lambda''$, it suffices to show that direction vectors for the two lines are not (real) scalar multiples of one another. But the direction vector for l_2'' is orthogonal to the radial vector for z_0, so the dot product calculation carried out above immediately implies the desired result.

Before beginning work on the second arc endpoint, $z_1 = \lambda'$, we introduce another piece of notation. For $0 \le t \le 1$, let z_t denote the point on the boundary arc corresponding to t; this simply extends our usage of z_0 and z_1.

We first establish that there is an $\varepsilon > 0$ such that for all $t \in (1-\varepsilon, 1)$, $(\partial f/\partial x)(z_t, t) \neq 0$. To this end, note first that neither z_t nor $z_t^p - t$ can be equal to 0 for $t \in (0,1)$: from (2.3), $z_t^p - t = 0$ would imply that $(1-t)^r = 0$, clearly an impossibility for all $t \in (0,1)$; since $q > 0$, the same thing rules out $z_t = 0$ for the relevant values of t.

Now suppose, for the sake of contradiction, that there is a sequence $\{t_n\}_1^\infty \subseteq (0,1)$ converging to 1 such that $(\partial f/\partial x)(z_{t_n}, t_n) = 0$ for all n. But

$$\begin{aligned}
\frac{\partial f}{\partial x}(z_{t_n}, t_n) &= q z_{t_n}^{q-1}(z_{t_n}^p - t_n)^r + pr z_{t_n}^{p+q-1}(z_{t_n}^p - t_n)^{r-1} \\
&= z_{t_n}^{q-1}(z_{t_n}^p - t_n)^{r-1}[q(z_{t_n}^p - t_n) + pr z_{t_n}^p] \\
&= z_{t_n}^{q-1}(z_{t_n}^p - t_n)^{r-1}[(pr+q)z_{t_n}^p - q t_n]
\end{aligned}$$

for all n by (2.10), so the observations in the previous paragraph allow us to conclude that

$$(pr+q)z_{t_n}^p - q t_n = 0 \tag{2.12}$$

for all n. Since $t_n \to 1$ as $n \to \infty$, $z_{t_n} \to z_1$ as well. Therefore, taking limits in (2.12) as $n \to \infty$ yields the identity $(pr+q)z_1^p = q$. As p, q, and r are all positive, however,

$$\left|(pr+q)z_1^p\right| = |pr+q| = pr+q > q = |q|,$$

giving us a contradiction. Thus $(\partial f/\partial x)(z_t, t) \neq 0$ for all t sufficiently close to 1.

Now from Lemma 2.1, the above implies that the complex-valued function $t \mapsto z_t$ is differentiable for all t sufficiently close to 1 with derivative

$$v_t = -\frac{\overline{w}_{1,t} w_{2,t}}{|w_{1,t}|^2} = -w_{2,t}/w_{1,t},$$

where $w_{1,t} = (\partial f/\partial x)(z_t, t)$ and $w_{2,t} = (\partial f/\partial t)(z_t, t)$. To show that the boundary arc has a well-defined tangent line at $z_1 = \lambda'$, then, it suffices to show that $\lim_{t\to 1} v_t$ exists and is nonzero. We begin with the former.

To ease notation slightly, we will henceforth let u_t denote $z_t^p - t$; so, to rephrase something we established above, z_t and u_t are both nonzero for the values of t with which we are concerned. Using the expressions for $\partial f/\partial x$ and $\partial f/\partial t$ from (2.10) and (2.11), as well as the identity $z_t^q u_t^r = (1-t)^r$ from (2.3), we obtain as an expression for v_t

$$v_t = \frac{1 - z_t^p}{1-t} \frac{r z_t}{(pr+q) z_t^p - qt}.$$

Examining different pieces of this expression, first notice that $r z_t \to r z_1$ as $t \to 1$. Since $z_1^q u_1^r = (1-1)^r = 0$, it follows that $u_1 = 0$, i.e. that $z_1^p = 1$; therefore $(pr+q) z_t^p - qt \to (pr+q) z_1^p - q = pr$ as $t \to 1$. Regarding the remaining fraction in the expression for v_t, observe that

$$\frac{1 - z_t^p}{1-t} = 1 - \frac{u_t}{1-t}.$$

But

$$\left(\frac{u_t}{1-t}\right)^r = z_t^{-q},$$

so a continuity argument shows that as $t \to 1$, $u_t/(1-t)$ approaches some rth root of z_1^{-q}, call it α. Putting all of the above together, we see that $\lim_{t\to 1} v_t$ exists, call it v, and that

$$v = (1-\alpha)\frac{r z_1}{pr} = \frac{(1-\alpha) z_1}{p}.$$

As in the method used at the first arc endpoint z_1, to complete the proof it suffices to establish that the dot product of v with the radial vector for z_1 is nonzero. Using the expression for such dot products derived earlier, we see that it is equal to

$$\operatorname{Re}\left(\frac{(1-\alpha) z_1}{p} \overline{z}_1\right) = \operatorname{Re}\left(\frac{1-\alpha}{p}\right) = \frac{1 - \operatorname{Re}\alpha}{p}. \tag{2.13}$$

Since α is an rth root of z_1^{-q}, certainly α lies on the unit circle. From (2.13), then, it suffices to show that $\alpha \neq 1$. Reducing the problem one additional time, notice that α cannot be equal to 1 if $z_1^{-q} \neq 1$. We thus focus on proving this last statement.

We begin by showing that $b' \nmid q$. This is evident if $b'' \leq b'$, since $0 < q < b''$. Hence suppose that $b'' > b'$ and that $b' \mid q$, say $q = kb'$. As we are assuming that $b' > 1$, it follows that $b'' \geq 3$. Additionally, $a'' < b'' - 1$ and $a'' \neq b''/2$. For the former, if $a'' = b'' - 1$, then we would have that $r_1 = b''$, $r_2 = a'' = b'' - 1 \geq 2$, and $r_3 = 1 = r_m$; but then m is odd, contradicting one of the hypotheses for the boundary arc to be described by (2.3). For the latter, if $a'' = b''/2$, then $a''/b'' = 1/2$, so in fact $a'' = 1$ and $b'' = 2$; but this contradicts the fact that $a'' > 1$. We now consider two cases:

- If $a'' < b''/2$, then, since $r_2 = a'' > 1$, we know that $r \leq r_2 = a'' < b''/2$.
- If $a'' > b''/2$, then $r_3 = b'' - a''$. But $1 = b'' - (b'' - 1) < b'' - a''$ and $b'' - a'' < b'' - b''/2 = b''/2$, so it follows that $r \leq r_3 = b'' - a'' < b''/2$.

Hence $r < b''/2$ regardless. Now, a'/b' and a''/b'' are consecutive terms in the Farey series $\mathcal{F}_n$, so $a''b' - a'b'' = 1$ and $a''b' \equiv_{b''} 1$. Therefore

$$-r \equiv a''q = a''(kb') = (a''b')k \equiv k \pmod{b''},$$

so $k = jb'' - r$ for some $j \geq 1$. Noting again that $b' > 1$, we obtain that

$$b'' = (b''/2) \cdot 2 < (b'' - r)b' \leq (jb'' - r)b' = kb' = q < b''.$$

This contradiction at last allows us to conclude that $b' \nmid q$ when $b'' > b'$, so $b' \nmid q$ in all cases.

Returning to the larger question, observe that $z_1^{-q} = e^{2\pi i(-a'q/b')}$. Since a' and b' are relatively prime (with $a' > 0$) and $b' \nmid q$, we can conclude that b' does not divide $a'q$. Thus $a'q/b'$ is not an integer, whence $z_1^{-q} \neq 1$. This completes the proof. $\square$

Theorem 2.4 (Krieger-Murcko). *Fix an $n > 3$. Suppose that $\lambda' = e^{2\pi i(a'/b')}$ and $\lambda'' = e^{2\pi i(a''/b'')}$ are consecutive roots of unity as in Theorem 2.2 and that the boundary arc of Θ_n connecting λ' and λ'' is described by (2.6). Then at the endpoints λ' and λ'' of the boundary arc, the arc has well-defined tangent lines l_1' and l_1'', respectively. Also, if l_2' and l_2'' denote the lines tangent to the unit circle $|z| = 1$ at λ' and λ'', respectively, then l_1' makes a nonzero angle with l_2' at λ', and likewise for l_1'' and l_2'' at λ''.*

Proof. To begin, note that we may assume that $a'/b' \neq 0/1$, for the associated boundary arc was addressed at the beginning of the proof of Theorem 2.3. In particular, this implies that $a' > 0$ and $b' > 1$.

Now let $f(z,t) = (z^b - t)^d - (1-t)^d z^q$, $z_0 = e^{2\pi i(a'/b')}$, and $z_1 = e^{2\pi i(a''/b'')}$. Observe that

$$f(z_1, 1) = (z_1^b - 1)^d - (1-1)^d z_1^q = \left((e^{2\pi i(a''/b'')})^b - 1\right)^d = (e^{2\pi i a''} - 1)^d = 0,$$

where we have made use of the fact that $b = b''$. Moreover, since $0 < q < b'' = b$ and $bd - q \equiv_b -q$, it follows that $b \nmid (bd-q)$. As a'' and b'' are relatively prime with $a'' \neq 0$ (this holds because $b'' > 1$—see the statement of Theorem 2.2), we additionally obtain $b \nmid a''(bd-q)$. Thus $z_1^{bd-q} = e^{2\pi i(a''(bd-q)/b'')} \neq 1$ and

$$f(z_1, 0) = z_1^{bd} - z_1^q = z_1^{bd-q} z_1^q - z_1^q \neq 1 \cdot z_1^q - z_1^q = 0;$$

hence it must be the case that $f(z_0, 0) = 0$.

To complete the preliminaries, we have

$$\frac{\partial f}{\partial x} = bdz^{b-1}(z^b - t)^{d-1} - qz^{q-1}(1-t)^d$$

and

$$\frac{\partial f}{\partial t} = dz^q(1-t)^{d-1} - d(z^b - t)^{d-1}.$$

For the endpoint z_0, we first calculate

$$\frac{\partial f}{\partial x}(z_0, 0) = bdz_0^{b-1} z_0^{b(d-1)} - qz_0^{q-1} = bdz_0^{bd-1} - qz_0^{q-1} = (bd-q)z_0^{q-1};$$

note that this uses the identity $z_0^{bd} - z_0^q = f(z_0, 0) = 0$. But $bd - q \geq b - q > b - b = 0$, whence $\partial f/\partial x$ is nonzero at $(z_0, 0)$. We now proceed exactly as we did with the arc endpoint z_0 in the previous proof. Letting $w_1 = (bd-q)z_0^{q-1}$ and

$$w_2 = \frac{\partial f}{\partial t}(z_0, 0) = dz_0^q - dz_0^{b(d-1)} = d(z_0^q - z_0^{bd-b}) = dz_0^q(1 - z_0^{-b}),$$

it suffices to show that the dot product of $v = \overline{w}_1 w_2 / |w_1|^2$ with the radial vector for z_0 is nonzero. As

$$v\overline{z}_0 = -\frac{\overline{w}_1 w_2}{|w_1|^2}\overline{z}_0 = -\frac{\left[(bd-q)\overline{z}_0^{q-1}\right]\left[dz_0^q(1 - z_0^{-b})\right]}{(bd-q)^2}\overline{z}_0 = -\frac{d(1 - z_0^{-b})}{bd-q},$$

the desired dot product is equal to

$$\mathrm{Re}\,(v\overline{z}_0) = -\frac{d}{bd-q}\big(1 - \mathrm{Re}\,(z_0^{-b})\big). \tag{2.14}$$

Since $a''b' - a'b'' = 1$ (this again follows from viewing a'/b' and a''/b'' as consecutive terms in the Farey series $\mathcal{F}_n$), we have that

$$z_0^{-b} = e^{2\pi i(-a'b''/b')} = e^{2\pi i(-(a''b'-1)/b')} = e^{-2\pi i a''}e^{2\pi i/b'} = e^{2\pi i/b'},$$

whence z_0^{-b} lies the unit circle but is not equal to 1 (recall that $b' > 1$). Thus $\mathrm{Re}\,(z_0^{-b}) < 1$, and since $d \geq 1$ we can conclude that the dot product in (2.14) is nonzero.

For the endpoint z_1, we continue in the mold of the proof of Theorem 2.3. That is, we begin by establishing that there is an $\varepsilon > 0$ such that $(\partial f/\partial x)(z_t, t) \neq 0$ for all $t \in (1-\varepsilon, 1)$. To this end, we first have that neither z_t nor $u_t = z_t^b - t$ can be equal to 0 for all $t \in (0,1)$: for $z_t = 0$ implies

$$0 = f(z_t, t) = f(0, t) = (-t)^d,$$

while $u_t = 0$ implies

$$0 = f(z_t, t) = u_t^d - (1-t)^d z_t^q = -(1-t)^d z_t^q.$$

Now suppose, for the sake of contradiction, that there is a sequence $\{t_n\}_1^\infty \subseteq (0,1)$ converging to 1 such that $(\partial f/\partial x)(z_{t_n}, t_n) = 0$ for all n. But using the identity $u_t^d = (1-t)^d z_t^q$, we have that

$$\begin{aligned} z_t \frac{\partial f}{\partial x}(z_t, t) &= bdz_t^b u_t^{d-1} - qz_t^q(1-t)^d = bdz_t^b u_t^{d-1} - qu_t^d \\ &= \big(bdz_t^b - q(z_t^b - t)\big)u_t^{d-1} = \big((bd-q)z_t^b + qt\big)u_t^{d-1}, \end{aligned}$$

so that by the observations in the previous paragraph $(bd-q)z_{t_n}^b + qt_n = 0$ for all n. Since $t_n \to 1$ and $z_{t_n} \to z_1$ as $n \to \infty$, this implies that

$$0 = \lim_{n\to\infty}\big((bd-q)z_{t_n}^b + qt_n\big) = (bd-q)z_1^b + q = bd,$$

clearly a contradiction since b and d are both positive. Thus we can conclude that $(\partial f/\partial x)(z_t, t)$ is nonzero for all t sufficiently close to 1.

Now take $w_{1,t} = (\partial f/\partial x)(z_t, t)$ and $w_{2,t} = (\partial f/\partial t)(z_t, t)$, and let

$$v_t = -\frac{\overline{w}_{1,t}w_{2,t}}{|w_{1,t}|^2} = -w_{2,t}/w_{1,t},$$

so that v_t is a tangent vector to the boundary arc for all t close to 1; as in the previous proof, we will show that $v = \lim_{t\to 1} v_t$ exists and that the dot product of this vector with the radial vector for z_1 is nonzero.

Using the expressions for $\partial f/\partial x$ and $\partial f/\partial t$ together with the identity $u_t^d = (1-t)^d z_t^q$, we see that

$$\begin{aligned} v_t &= -\frac{w_{2,t}}{w_{1,t}} = -\frac{u_t w_{2,t}}{u_t w_{1,t}} = -\frac{d u_t z_t^q (1-t)^{d-1} - d z_t^q (1-t)^d}{b d z_t^{b+q-1}(1-t)^d - q u_t z_t^{q-1}(1-t)^d} \\ &= -\frac{d u_t z_t - d(1-t) z_t}{b d z_t^b (1-t) - q u_t (1-t)} = -\frac{u_t - 1 + t}{1-t} \frac{d z_t}{b d z_t^b - q u_t} \\ &= -\frac{z_t^b - 1}{1-t} \frac{d z_t}{(bd-q) z_t^b + qt}. \end{aligned}$$

Examining pieces of this final expression, we find first that

$$\lim_{t\to 1} \frac{d z_t}{(bd-q) z_t^b + qt} = \frac{d z_1}{(bd-q) z_1^b + q} = \frac{d z_1}{bd - q + q} = z_1/b.$$

In addition,

$$-\frac{z_t^b - 1}{1-t} = 1 - \frac{u_t}{1-t}.$$

The fact that $(u_t/(1-t))^d = z_t^q$ for $t \neq 1$ implies that $u_t/(1-t)$ approaches some dth root of z_1^q as $t \to 1$. Let α denote this limit. Since $a''q \equiv_{b''} -1$ and $b'' > 1$, we see that $z_1^q = e^{2\pi i (a''q/b'')} = e^{-2\pi i/b''} \neq 1$. Thus α lies on the unit circle but is not equal to 1; in particular, $\operatorname{Re}\alpha < 1$.

Putting the above together, we obtain that

$$v = \lim_{t\to 1} v_t = (1-\alpha) z_1/b,$$

whence the dot product of v and the radial vector for z_1 is

$$\operatorname{Re}(v\bar{z}_1) = \operatorname{Re}\left(\frac{(1-\alpha)z_1}{b}\bar{z}_1\right) = \frac{\operatorname{Re}(1-\alpha)}{b} = \frac{1-\operatorname{Re}\alpha}{b} > \frac{1-1}{b} = 0.$$

This completes the proof. $\square$

Chapter 3

Analytic Perturbations

In this chapter we review some important results from perturbation theory for linear operators on the finite-dimensional vector space $\mathbb{C}^n$ (i.e. $n \times n$ matrices with coefficients in $\mathbb{C}$). Most of the material is drawn from Chapter 2 of [6]. In all of what follows, $M_n(\mathbb{C})$ denotes the set of complex $n \times n$ matrices.

3.1 Perturbed Eigenvalues and Eigenprojections

We begin with a definition to place us on solid ground.

Definition 3.1. *An* analytic perturbation *of a matrix $T_0 \in M_n(\mathbb{C})$ is a power series*

$$T(\varepsilon) = T_0 + A(\varepsilon) = T_0 + \varepsilon A_1 + \varepsilon^2 A_2 + \cdots$$

in which the coefficients $A_1, A_2, \ldots$ are all elements of $M_n(\mathbb{C})$ as well. We refer to $T(\varepsilon)$ as an analytically perturbed matrix.

Such a power series will have a radius of convergence $r_0 \in [0, \infty]$ just as does a standard power series in which the coefficients are complex numbers rather than matrices. That is, $T(\varepsilon)$ will converge for all complex ε satisfying $|\varepsilon| < r_0$ and diverge for all ε with $|\varepsilon| > r_0$; the value of r_0 depends on the entries in $A_1, A_2, \ldots$ Henceforth we assume that any analytically perturbed matrix with which we are concerned has positive radius of convergence r_0.

If $T(\varepsilon)$ is an analytic perturbation of T_0, $T(\varepsilon)$ has a fixed number of distinct eigenvalues except at certain "exceptional" values of ε, only a finite number of which lie in any compact set. If λ is an eigenvalue of T_0, then

$T(\varepsilon)$ possesses a collection of associated perturbed eigenvalues, call them $\lambda_1(\varepsilon), \lambda_2(\varepsilon), \ldots, \lambda_k(\varepsilon)$, each of which can be represented in a power series–like form called a *Puiseux series*:

$$\lambda_j(\varepsilon) = \lambda + c_{1,j}\varepsilon^{1/p_j} + c_{2,j}\varepsilon^{2/p_j} + \cdots , \tag{3.1}$$

where p_j is a positive integer. We refer to $\lambda_1(\varepsilon), \ldots, \lambda_k(\varepsilon)$ as the λ-group eigenvalues for $T(\varepsilon)$, since they all converge to λ as $\varepsilon \to 0$.

Recall from Chapter 2 that the *multiplicity* of an eigenvalue refers to its algebraic multiplicity, or equivalently the dimension of its generalized eigenspace. It is known that at any non-exceptional value of ε as described above, the sum of the multiplicities of the individual λ-group eigenvalues in $T(\varepsilon)$ is equal to the multiplicity of λ in T_0.

There are also useful results relating to eigenprojection matrices. Certainly at non-exceptional values of ε, it makes sense to talk about eigenprojections for the individual perturbed eigenvalues $\lambda_j(\varepsilon)$. If we choose $r > 0$ small enough that the domain $D = \{\, \varepsilon \mid |\varepsilon| < r \,\}$ contains no exceptional points except possibly 0, we can further define a matrix-valued function $P^*(\lambda, \varepsilon)$ on D equal to the sum of these individual eigenprojections. As it turns out $P^*(\lambda, \varepsilon)$ can be expressed as an analytic perturbation of $P^*(\lambda)$, the eigenprojection for T_0 associated with λ. We refer to $P^*(\lambda, \varepsilon)$ as the *total projection* for the λ-group. If the multiplicity of λ in T_0 is m, then $P^*(\lambda, \varepsilon)$ projects onto an m-dimensional subspace of $\mathbb{C}^n$. This subspace, which generally depends on ε, is denoted $M(\lambda, \varepsilon)$; we refer to it as the *perturbed eigenspace* for the λ-group. For the remainder of the chapter, we only concern ourselves with $T(\varepsilon)$ on a domain as described above.

We illustrate the concepts introduced in this section with an example.

Example 3.1. Consider the analytically perturbed matrix

$$T(\varepsilon) = \begin{bmatrix} 2 - \varepsilon + \varepsilon^2 & 3\varepsilon^2 \\ \varepsilon^2 & 2 - \varepsilon - \varepsilon^2 \end{bmatrix}.$$

To be explicit, we can write this as

$$\begin{aligned} T(\varepsilon) &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} + \begin{bmatrix} -\varepsilon + \varepsilon^2 & 3\varepsilon^2 \\ \varepsilon^2 & -\varepsilon - \varepsilon^2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} + \varepsilon \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} + \varepsilon^2 \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix}. \end{aligned}$$

Therefore

$$T_0 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad A_1 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix};$$

also,

$$A(\varepsilon) = \varepsilon A_1 + \varepsilon^2 A_2 = \begin{bmatrix} -\varepsilon + \varepsilon^2 & 3\varepsilon^2 \\ \varepsilon^2 & -\varepsilon - \varepsilon^2 \end{bmatrix}.$$

Evidently the radius of convergence of $T(\varepsilon)$ is ∞.

T_0 has a single eigenvalue of multiplicity 2, namely 2. It is straightforward to see, then, that the eigenprojection for T_0 associated with the eigenvalue 2 is the identity matrix I.

Puiseux series for the perturbed eigenvalues in the 2-group of $T(\varepsilon)$ can be determined straightforwardly in this example by calculating the characteristic polynomial of $T(\varepsilon)$ and factoring it. This process yields $\lambda_1(\varepsilon) = 2 - \varepsilon + 2\varepsilon^2$ and $\lambda_2(\varepsilon) = 2 - \varepsilon - 2\varepsilon^2$. Since these Puiseux series are simply polynomials in ε, there is no restriction on the values we may assign the positive integers p_1 and p_2 in (3.1). From the expressions for the perturbed eigenvalues, we see that 0 is the only exceptional point; that is, $\lambda_1(\varepsilon) = \lambda_2(\varepsilon)$ if and only if $\varepsilon = 0$.

With a bit more effort, the individual eigenprojections for $\lambda_1(\varepsilon)$ and $\lambda_2(\varepsilon)$, valid for all $\varepsilon \neq 0$, can be calculated as well. They are

$$P^*\big(\lambda_1(\varepsilon)\big) = \begin{bmatrix} 3/4 & 3/4 \\ 1/4 & 1/4 \end{bmatrix} \quad \text{and} \quad P^*\big(\lambda_2(\varepsilon)\big) = \begin{bmatrix} 1/4 & -3/4 \\ -1/4 & 3/4 \end{bmatrix}.$$

From this we see that $P^*(2,\varepsilon) = P^*\big(\lambda_1(\varepsilon)\big) + P^*\big(\lambda_2(\varepsilon)\big)$, the total projection for the 2-group, is equal to the identity matrix; that is, the perturbation term for the total projection is equal to 0 in this case. This is to be expected from the comment in Section 2.1 regarding sums of eigenprojections; it also fits with the earlier statement in this section implying that $P^*(2,\varepsilon)$ is an analytic perturbation of the eigenprojection for T_0 corresponding to the eigenvalue 2. Since $P^*(2,\varepsilon) = I$, it is evident that $M(2,\varepsilon) = \mathbb{C}^2$.

This example was rather contrived insofar as we could easily determine complete expressions for the Puiseux series, the individual eigenprojections, and the total projection. In the general case these objects can only be approximated; for example, each term in the series expansion for the total projection must be calculated using a fairly complicated formula (see [6], p. 77). For the perturbed eigenvalues, which are of significant interest for analyzing the hybrid Cesaro limits, at least a portion of the Puiseux series can be obtained by a type of reduction process.

3.2 The Reduction Process

We begin here with a continuation of the previous example.

Example 3.2. Let

$$T(\varepsilon) = \begin{bmatrix} 2-\varepsilon+\varepsilon^2 & 3\varepsilon^2 \\ \varepsilon^2 & 2-\varepsilon-\varepsilon^2 \end{bmatrix}.$$

Now consider the matrix

$$\widetilde{T}(\varepsilon) = \frac{1}{\varepsilon}(T(\varepsilon)-2I)P^*(2,\varepsilon) = \frac{1}{\varepsilon}(T(\varepsilon)-2I)I = \begin{bmatrix} -1+\varepsilon & 3\varepsilon \\ \varepsilon & -1-\varepsilon \end{bmatrix}.$$

Observe that $\widetilde{T}(\varepsilon)$ is actually an analytic perturbation of the matrix

$$\widetilde{T}_0 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Clearly $\widetilde{T}_0$ has -1 as an eigenvalue of multiplicity 2; moreover, the perturbed -1-group eigenvalues can be directly calculated in the same way as in Example 3.1. They are $\tilde{\lambda}_1(\varepsilon) = -1+2\varepsilon$ and $\tilde{\lambda}_2(\varepsilon) = -1-2\varepsilon$. Observe the relationship between the $\lambda_j(\varepsilon)$, as determined in Example 3.1, and the $\tilde{\lambda}_j(\varepsilon)$: in each case,

$$\lambda_j(\varepsilon) = 2+\varepsilon\tilde{\lambda}_j(\varepsilon).$$

In other words, the new perturbed eigenvalues $\tilde{\lambda}_j(\varepsilon)$ are obtained from the old ones by subtracting off the unperturbed part, 2, and dividing by ε. In particular, the first-order coefficients of the Puiseux series for $\lambda_1(\varepsilon) = 2-\varepsilon+2\varepsilon^2$ and $\lambda_2(\varepsilon) = 2-\varepsilon-2\varepsilon^2$, namely -1 for both cases, are precisely the eigenvalues of the unperturbed matrix $\widetilde{T}_0$.

The above is an instance of the *reduction process*. More specifically, the reduction process consists of determining $\widetilde{T}(\varepsilon)$ from $T(\varepsilon)$ in order to gain information about the Puiseux series of the perturbed eigenvalues of $T(\varepsilon)$. As alluded to at the end of the previous section, this is primarily useful when we cannot immediately generate these Puiseux series, as we could in the previous examples by factoring the characteristic polynomial.

For the remainder of the section, we assume that $T(\varepsilon) = T_0 + A(\varepsilon) = T_0 + \varepsilon A_1 + \varepsilon^2 A_2 + \cdots$ is an $n \times n$ analytically perturbed matrix and that λ is a semisimple eigenvalue of T_0 of multiplicity m. In the general setting, the reduction process works as follows. We *reduce* $T(\varepsilon)$ for λ to the matrix

$$\widetilde{T}(\varepsilon) = \frac{1}{\varepsilon}(T(\varepsilon)-\lambda I)P^*(\lambda,\varepsilon). \tag{3.2}$$

$\widetilde{T}(\varepsilon)$ is necessarily an analytically perturbed matrix itself. To see this, first note that the product of two analytically perturbed matrices is again an analytically perturbed matrix. Therefore $(T(\varepsilon) - \lambda I)P^*(\lambda, \varepsilon)$ is certainly an analytically perturbed matrix. The "constant" term, $(T_0 - \lambda I)P^*(\lambda, 0)$, acts on $\mathbb{C}^n$ by projecting a vector onto the generalized eigenspace for λ in T_0, and then multiplying by $T_0 - \lambda I$. This is identical to the action of the 0 matrix precisely when the eigenspace and generalized eigenspace for λ in T_0 are the same, or equivalently when λ is semisimple for T_0. In other words, the constant term $(T_0 - \lambda I)P^*(\lambda, 0)$ is equal to 0 precisely when λ is a semisimple eigenvalue of T_0. Since we assume λ is semisimple above, $(T(\varepsilon) - \lambda I)P^*(\lambda, \varepsilon)$ will have the form $0 + \varepsilon B_1 + \varepsilon^2 B_2 + \cdots$ for some matrices $B_1, B_2, \ldots$ Dividing by ε therefore yields the analytically perturbed matrix $B_1 + \varepsilon B_2 + \cdots$. One can show that $\widetilde{T}(0) = B_1 = P^*(\lambda, 0)A_1 P^*(\lambda, 0)$.

Returning to the question of how the reduction process affects the eigenvalues of the matrices in question, we begin by writing $\mathbb{C}^n = M(\lambda, \varepsilon) \oplus M'(\varepsilon)$, where

$$M'(\varepsilon) = \bigoplus_{\mu \neq \lambda} M(\mu, \varepsilon),$$

the direct sum being taken over all eigenvalues μ of T_0 not equal to λ. Since $M(\lambda, \varepsilon)$ has dimension m, $M'(\varepsilon)$ has dimension $n - m$.

If $v \in M'(\varepsilon)$ for a fixed ε, then $P^*(\lambda, \varepsilon)v = 0$, whence $\widetilde{T}(\varepsilon)v = 0$ as well. In other words, $\widetilde{T}(\varepsilon)$ maps all of $M'(\varepsilon)$ to 0, so $M'(\varepsilon)$ is contained in the (generalized) eigenspace for 0. Since the direct sum of the generalized eigenspaces for $\widetilde{T}(\varepsilon)$ is equal to $\mathbb{C}^n$, there must exist a collection $\widetilde{\lambda}_1(\varepsilon), \ldots, \widetilde{\lambda}_k(\varepsilon)$ of perturbed eigenvalues of $\widetilde{T}(\varepsilon)$, each of which has a generalized eigenvector that does not lie in $M'(\varepsilon)$. More specifically, there must exist linearly independent vectors $v_1(\varepsilon), \ldots, v_m(\varepsilon)$ such that

(1) $v_i(\varepsilon)$ is a generalized eigenvector of some $\widetilde{\lambda}_j(\varepsilon)$ for each i;

(2) if $W(\varepsilon) = \operatorname{span}\{v_1(\varepsilon), \ldots, v_m(\varepsilon)\}$, then $\mathbb{C}^n = W(\varepsilon) \oplus M'(\varepsilon)$.

Observe that condition (2) implies that $v_i(\varepsilon) \notin M'(\varepsilon)$ for each i. Therefore $v_i'(\varepsilon) = P^*(\lambda, \varepsilon)v_i(\varepsilon)$, the component of $v_i(\varepsilon)$ in $M(\lambda, \varepsilon)$, is always nonzero. One can now show that the $v_i'(\varepsilon)$ are all linearly independent, so that they span $M(\lambda, \varepsilon)$. Further, the $v_i'(\varepsilon)$ are all still generalized eigenvectors of the $\widetilde{\lambda}_j(\varepsilon)$: if $\left[\widetilde{T}(\varepsilon) - \lambda_j(\varepsilon)I\right]^l v_i(\varepsilon) = 0$, then $\left[\widetilde{T}(\varepsilon) - \lambda_j(\varepsilon)I\right]^l v_i'(\varepsilon) = 0$ as well. Using the fact that $v_i'(\varepsilon) \in M(\lambda, \varepsilon)$, it also follows that

$$\left[T(\varepsilon) - \left(\lambda + \varepsilon\widetilde{\lambda}_j(\varepsilon)\right)I\right]^l v_i'(\varepsilon) = 0.$$

In other words, $v_i'(\varepsilon)$ is a generalized eigenvector of $T(\varepsilon)$ associated with the perturbed eigenvalue $\lambda + \varepsilon\widetilde{\lambda}_j(\varepsilon)$.

From all of this, we see that $\lambda + \varepsilon\widetilde{\lambda}_1(\varepsilon), \ldots, \lambda + \varepsilon\widetilde{\lambda}_k(\varepsilon)$ must be precisely the λ-group eigenvalues of $T(\varepsilon)$. And conversely, if $\lambda(\varepsilon)$ is a λ-group eigenvalue of $T(\varepsilon)$, then

$$\frac{1}{\varepsilon}(\lambda(\varepsilon) - \lambda)$$

is a perturbed eigenvalue of $\widetilde{T}(\varepsilon)$. Most importantly, the Puiseux series of each λ-group eigenvalue of $T(\varepsilon)$ must begin $\lambda_j(\varepsilon) = \lambda + \varepsilon\widetilde{\lambda}_j(0) + \cdots$ for some j; in other words, these Puiseux series involve no fractional powers of ε less than 1. So by calculating the values of the $\widetilde{\lambda}_j(0)$—they are all eigenvalues of $\widetilde{T}(0)$—we obtain the first-order coefficients in the Puiseux series of the λ-group eigenvalues of $T(\varepsilon)$.

A subtle but rather important point in the discussion above is that $\widetilde{T}(\varepsilon)$ may be viewed, for the purposes of obtaining eigenvalue information, as acting not on all of $\mathbb{C}^n$ but just on the perturbed eigenspace $M(\lambda, \varepsilon)$. Certainly $\widetilde{T}(\varepsilon)$ maps $M(\lambda, \varepsilon)$ to itself, and further all the reduced λ-group eigenvalues of $T(\varepsilon)$ (that is, the $\widetilde{\lambda}_j(\varepsilon)$) occur as eigenvalues of $\widetilde{T}(\varepsilon)$ within the subspace $M(\lambda, \varepsilon)$. We therefore lose nothing by ignoring the trivial action of $\widetilde{T}(\varepsilon)$ on $M'(\varepsilon)$. In fact, we gain a bit of clarity, for we would rather ignore the eigenvalue 0 of $\widetilde{T}(\varepsilon)$ as it relates to the subspace $M'(\varepsilon)$. It may happen, in particular, that 0 is also an eigenvalue of $\widetilde{T}(\varepsilon)$ in $M(\lambda, \varepsilon)$, and it is preferable not to group this appearance of the eigenvalue 0 with its appearance for $M'(\varepsilon)$.

Using this idea, we can describe how the reduction process may be performed iteratively to yield successively more information about the Puiseux series coefficients for the perturbed eigenvalues. Given $T_1(\varepsilon)$, if λ_1 is a semisimple eigenvalue of $T_1(0)$, we first reduce $T_1(\varepsilon)$ for λ_1 to the matrix $T_2(\varepsilon)$. Viewing the reduced matrix as acting on the subspace $M_1(\lambda_1, \varepsilon) = M(\lambda_1, \varepsilon)$ as above, we can ask whether $T_2(0)$ has any semisimple eigenvalues in $M_1(\lambda_1, 0)$. If so—say λ_2 is such an eigenvalue—we reduce $T_2(\varepsilon)$ for λ_2 to the matrix

$$T_3(\varepsilon) = \frac{1}{\varepsilon}\big(T_2(\varepsilon) - \lambda_2 P^*(\lambda_1, 0)\big)\widetilde{P}^*(\lambda_2, \varepsilon). \tag{3.3}$$

This reduction equation takes a little more explanation. To begin, $\widetilde{P}^*(\lambda_2, \varepsilon)$ is the projection onto the perturbed eigenspace for the λ_2-group *in* $M_1(\lambda_1, \varepsilon)$. If $\lambda_2 = 0$, this may not be equal to the total projection for the λ_2-group, since the total projection includes the subspace described above that we wish to ignore. Nonetheless, it is still guaranteed that $\widetilde{P}^*(\lambda_2, \varepsilon)$ is an analytically perturbed matrix. Thus we refer to it as an *analytic projection*.

Continuing, $P^*(\lambda_1, 0)$ takes the place of the identity matrix between (3.2) and (3.3). To maintain the perspective of only paying attention to $M_1(\lambda_1, \varepsilon)$, it is more appropriate to use $P^*(\lambda_1, 0)$, which is the identity operator on $M_1(\lambda_1, 0)$ but maps the subspace we wish to ignore to 0. This point, however, is less essential than the one made in the previous paragraph. We could actually retain I in (3.3) and obtain the same $T_3(\varepsilon)$, as $\widetilde{P}^*(\lambda_2, \varepsilon)$ is a sub-projection of $P^*(\lambda_1, \varepsilon)$.

We now view $T_3(\varepsilon)$ as acting on the range of the matrix $\widetilde{P}^*(\lambda_2, \varepsilon)$, which we denote by $M_2(\lambda_2, \varepsilon)$; this is analogous to $T_2(\varepsilon)$ acting on the range of the matrix $P^*(\lambda, \varepsilon)$, since the latter range is equal to $M_1(\lambda_1, \varepsilon)$. Note that $M_2(\lambda_2, \varepsilon)$ is equal to the intersection of $M_1(\lambda_1, \varepsilon)$ with the entire perturbed eigenspace for the λ_2-group in $T_2(\varepsilon)$. If now $\lambda_3(\varepsilon)$ is a perturbed eigenvalue of $T_3(\varepsilon)$ in $M_2(\lambda_2, \varepsilon)$, then in a manner analogous to what we saw earlier it follows that $\lambda_2 + \varepsilon\lambda_3(\varepsilon)$ is a perturbed eigenvalue of $T_2(\varepsilon)$ in $M_1(\lambda_1, \varepsilon)$. Therefore

$$\lambda_1 + \varepsilon\big(\lambda_2 + \varepsilon\lambda_3(\varepsilon)\big) = \lambda_1 + \varepsilon\lambda_2 + \varepsilon^2\lambda_3(0) + \cdots$$

is a perturbed eigenvalue in the λ_1-group of $T(\varepsilon)$. This process can be repeated whenever we can find semisimple eigenvalues following a given reduction. As long as it continues, the coefficients in the Puiseux series are attached to positive integral powers of ε.

3.3 Applications to Stochastic Matrices

As described in Chapter 1, we are primarily concerned with analytic perturbations of *stochastic* matrices. This motivates the following definition.

Definition 3.2. *An* analytically perturbed stochastic matrix $P(\varepsilon)$ *is an analytic perturbation of a stochastic matrix* P_0 *such that for all sufficiently small positive* ε, $P(\varepsilon)$ *remains stochastic.*

This definition ensures that the analytically perturbed matrices with which we concern ourselves have nice probabilistic interpretations. Also, we often drop the adverb "analytically" and just refer to $P(\varepsilon)$ as a perturbed stochastic matrix.

The fact that the unit-circle eigenvalues of a stochastic matrix are always semisimple permits us to perform the reduction process for a perturbed stochastic matrix $P(\varepsilon) = P_0 + A(\varepsilon)$ at least once for any such eigenvalue. In Chapter 4 we go into more detail on the reduction process for unit-circle eigenvalues of $P(\varepsilon)$.

Chapter 4

Cesaro Limit Results

Throughout this chapter, $P(\varepsilon) = P_0 + A(\varepsilon)$ is an $n \times n$ analytically perturbed stochastic matrix.

4.1 Decomposing the Cesaro Limit

Combining the information in Sections 2.1 and 3.1, we see that for a given $P(\varepsilon)$,

$$\sum_{\lambda} P^*(\lambda, \varepsilon) = I,$$

where the sum is taken over all eigenvalues λ of P_0. This type of decomposition of the identity matrix suggests decomposing the hybrid Cesaro limit expression as well:

$$\frac{1}{N}\sum_{k=1}^{N} P^k(\varepsilon) = \frac{1}{N}\sum_{k=1}^{N} P^k(\varepsilon)\sum_{\lambda} P^*(\lambda, \varepsilon) = \sum_{\lambda}\frac{1}{N}\sum_{k=1}^{N} P^k(\varepsilon)P^*(\lambda, \varepsilon).$$

Since P_0 has only finitely many eigenvalues, we can investigate the overall hybrid Cesaro limit by separately examining

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon)P^*(\lambda, \varepsilon)$$

for each eigenvalue λ of P_0. This is very similar to the approach used in [4], more of which we shall see shortly. Before proceeding, however, we look at a pair of examples to get a sense of what will come.

4.2 Examples

In each of the following two examples, drawn from [8], we analyze the limit

$$\lim_{\varepsilon\downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon)P^*(\lambda,\varepsilon)$$

when λ is an eigenvalue of P_0 on the unit circle other than 1 (in both cases, $\lambda = -1$). Each example possesses a property that one might expect to cause the limit not to exist; we show, however, that this is not the case. Recall from Section 2.1 that the *multiplicity* of an eigenvalue is its algebraic multiplicity, or equivalently the degree of its associated generalized eigenspace.

Example 4.1. For $0 < \varepsilon \le 1$, let

$$P(\varepsilon) = P_0 + \varepsilon A_1 = \begin{bmatrix} 0 & 1-\varepsilon & 0 & \varepsilon & 0 & 0 \\ 1-\varepsilon & 0 & \varepsilon & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-\varepsilon & 0 & \varepsilon \\ 0 & 0 & 1-\varepsilon & 0 & \varepsilon & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

The unperturbed matrix

$$P_0 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

has each of 1 and -1 as eigenvalues of multiplicity 3. The -1-group eigenvalues for $P(\varepsilon)$ are $\lambda_1(\varepsilon) = -1$ and $\lambda_2(\varepsilon) = -1+\varepsilon$. The former has multiplicity 1, and is thus necessarily semisimple, while the latter is not semisimple: its algebraic multiplicity is 2, but its geometric multiplicity is 1. The total projection $P^*(-1,\varepsilon)$ for the -1-group is the sum of the individual eigenprojections

$$P^*\big(\lambda_1(\varepsilon)\big) = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

and

$$P^*\big(\lambda_2(\varepsilon)\big) = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 1 \\ -1 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & -1 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In addition, the nilpotent matrix associated with $\lambda_2(\varepsilon)$ is

$$D(\varepsilon) = \varepsilon M = \frac{\varepsilon}{2}\begin{bmatrix} 0 & 0 & -1 & 1 & 1 & -1 \\ 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix};$$

its index of nilpotence is 2, since $D^2(\varepsilon) = 0$.

Next we show that

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(-1,\varepsilon) = 0,$$

regardless of how fast $N(\varepsilon) \uparrow \infty$ as $\varepsilon \downarrow 0$.

To this end, first note that

$$\begin{aligned} P^k(\varepsilon)P^*(-1,\varepsilon) &= P^k(\varepsilon)\big[P^*\big(\lambda_1(\varepsilon)\big) + P^*\big(\lambda_2(\varepsilon)\big)\big] \\ &= \lambda_1^k(\varepsilon)P^*\big(\lambda_1(\varepsilon)\big) + \lambda_2^k(\varepsilon)P^*\big(\lambda_2(\varepsilon)\big) + k\lambda_2^{k-1}(\varepsilon)D(\varepsilon) \\ &= (-1)^k P^*\big(\lambda_1(\varepsilon)\big) + (-1+\varepsilon)^k P^*\big(\lambda_2(\varepsilon)\big) \\ &\qquad + k\varepsilon(-1+\varepsilon)^{k-1}M. \end{aligned}$$

Hence

$$\begin{aligned} &\frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon)P^*(-1,\varepsilon) \\ &\qquad = \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} (-1)^k P^*\big(\lambda_1(\varepsilon)\big) + \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} (-1+\varepsilon)^k P^*\big(\lambda_2(\varepsilon)\big) \\ &\qquad\quad + \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} k\varepsilon(-1+\varepsilon)^{k-1}M. \end{aligned} \tag{4.1}$$

Now consider the three terms in (4.1) individually. To start,

$$\frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)}(-1)^k P^*\big(\lambda_1(\varepsilon)\big) = \frac{-1+(-1)^{N(\varepsilon)}}{2N(\varepsilon)}P^*\big(\lambda_1(\varepsilon)\big).$$

Since $P^*\big(\lambda_1(\varepsilon)\big)$ is a constant matrix and

$$\begin{aligned}\left|\frac{-1+(-1)^{N(\varepsilon)}}{2N(\varepsilon)}\right| &= \frac{\left|-1+(-1)^{N(\varepsilon)}\right|}{2N(\varepsilon)}\\ &\leq \frac{2}{2N(\varepsilon)} \to 0\end{aligned}$$

as $\varepsilon \downarrow 0$, the first term contributes nothing to the overall limit. The second term behaves in the same way because $|-1+\varepsilon| \leq 1$, $-1+\varepsilon$ is bounded away from 1 as $\varepsilon \downarrow 0$, and $P^*\big(\lambda_2(\varepsilon)\big)$ is also a constant matrix.

As for the third term in (4.1), we see that

$$\begin{aligned}&\frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)} k\varepsilon(-1+\varepsilon)^{k-1}M\\ &\qquad= \frac{1-(-1+\varepsilon)^{N(\varepsilon)}\big[1+(2-\varepsilon)N(\varepsilon)\big]}{(2-\varepsilon)^2N(\varepsilon)}\varepsilon M\\ &\qquad= \left[\frac{\varepsilon}{(2-\varepsilon)^2N(\varepsilon)} - \frac{\varepsilon(-1+\varepsilon)^{N(\varepsilon)}}{(2-\varepsilon)^2N(\varepsilon)} - \frac{\varepsilon(-1+\varepsilon)^{N(\varepsilon)}}{2-\varepsilon}\right]M. \qquad (4.2)\end{aligned}$$

The numerator of each bracketed term in (4.2) approaches 0 as $\varepsilon \downarrow 0$: in each case the magnitude of the numerator is $\leq \varepsilon$. On the other hand, the first two denominators become unbounded and the third denominator approaches a finite nonzero limit. It follow that the entire bracketed coefficient of M in (4.2) goes to 0, and since M is itself a constant matrix this implies that the third term from (4.1) tends to the zero matrix as $\varepsilon \downarrow 0$.

We showed that all three terms in (4.1) approach 0 as $\varepsilon \downarrow 0$, thus establishing that the overall limit is 0.

Example 4.2. For $0 < \varepsilon \leq \frac{1}{2}$, let

$$P(\varepsilon) = \begin{bmatrix} 0 & 1-2\varepsilon & \varepsilon & \varepsilon^2 & 0 & \varepsilon-\varepsilon^2\\ 1-2\varepsilon & 0 & \varepsilon^2 & \varepsilon & \varepsilon-\varepsilon^2 & 0\\ 0 & 0 & \varepsilon+\varepsilon^2-\varepsilon^3 & 1-\varepsilon-\varepsilon^2+\varepsilon^3 & 0 & 0\\ 0 & 0 & 1-\varepsilon-\varepsilon^2+\varepsilon^3 & \varepsilon+\varepsilon^2-\varepsilon^3 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

The unperturbed matrix P_0 is the same as that in Example 4.1, having 1 and -1 as eigenvalues of multiplicity 3. The -1-group eigenvalues of $P(\varepsilon)$ are $\lambda_1(\varepsilon) = -1$, $\lambda_2(\varepsilon) = -1 + 2\varepsilon$, and $\lambda_3(\varepsilon) = -1 + 2\varepsilon + 2\varepsilon^2 - 2\varepsilon^3$; each has multiplicity 1. The associated eigenprojections, which sum to the total projection $P^*(-1, \varepsilon)$ for the -1-group, are

$$P^*\big(\lambda_1(\varepsilon)\big) = \frac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 0 & 1-\varepsilon & -1+\varepsilon \\ 0 & 0 & 0 & 0 & -1+\varepsilon & 1-\varepsilon \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & 0 & -2 & 2 \end{bmatrix},$$

$$P^*\big(\lambda_2(\varepsilon)\big) = \frac{1}{4}\begin{bmatrix} 2 & -2 & -1/\varepsilon & 1/\varepsilon & -1+\varepsilon & 1-\varepsilon \\ -2 & 2 & 1/\varepsilon & -1/\varepsilon & 1-\varepsilon & -1+\varepsilon \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$P^*\big(\lambda_3(\varepsilon)\big) = \frac{1}{4}\begin{bmatrix} 0 & 0 & 1/\varepsilon & -1/\varepsilon & 0 & 0 \\ 0 & 0 & -1/\varepsilon & 1/\varepsilon & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & -2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Notice that $P^*\big(\lambda_2(\varepsilon)\big)$ and $P^*\big(\lambda_3(\varepsilon)\big)$ become unbounded as $\varepsilon \downarrow 0$ because of the entries that involve $1/\varepsilon$.

As in Example 4.1, we show that

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(-1, \varepsilon) = 0,$$

again irrespective of the rate at which $N(\varepsilon) \uparrow \infty$.

To begin, define new matrices

$$M_1(\varepsilon) = \frac{1}{4}\begin{bmatrix} 2 & -2 & 0 & 0 & -1+\varepsilon & 1-\varepsilon \\ -2 & 2 & 0 & 0 & 1-\varepsilon & -1+\varepsilon \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$M_2 = \frac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & -2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$M_3(\varepsilon) = \frac{1}{4}\begin{bmatrix} 0 & 0 & -1/\varepsilon & 1/\varepsilon & 0 & 0 \\ 0 & 0 & 1/\varepsilon & -1/\varepsilon & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

so that $P^*\big(\lambda_2(\varepsilon)\big) = M_1(\varepsilon) + M_3(\varepsilon)$ and $P^*\big(\lambda_3(\varepsilon)\big) = M_2 - M_3(\varepsilon)$. Then

$$\begin{aligned} P^k(\varepsilon)P^*(-1,\varepsilon) &= P^k(\varepsilon)\big[P^*\big(\lambda_1(\varepsilon)\big) + P^*\big(\lambda_2(\varepsilon)\big) + P^*\big(\lambda_3(\varepsilon)\big)\big] \\ &= \lambda_1^k(\varepsilon)P^*\big(\lambda_1(\varepsilon)\big) + \lambda_2^k(\varepsilon)P^*\big(\lambda_2(\varepsilon)\big) + \lambda_3^k(\varepsilon)P^*\big(\lambda_3(\varepsilon)\big) \\ &= \lambda_1^k(\varepsilon)P^*\big(\lambda_1(\varepsilon)\big) + \lambda_2^k(\varepsilon)M_1(\varepsilon) + \lambda_3^k(\varepsilon)M_2 \\ &\qquad + \big(\lambda_2^k(\varepsilon) - \lambda_3^k(\varepsilon)\big)M_3(\varepsilon), \end{aligned}$$

and thus

$$\begin{aligned} &\frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon)P^*(-1,\varepsilon) \\ &\quad = \frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)} \lambda_1^k(\varepsilon)P^*\big(\lambda_1(\varepsilon)\big) + \frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)} \lambda_2^k(\varepsilon)M_1(\varepsilon) \\ &\qquad + \frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)} \lambda_3^k(\varepsilon)M_2 + \frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)} \big(\lambda_2^k(\varepsilon) - \lambda_3^k(\varepsilon)\big)M_3(\varepsilon). \end{aligned} \tag{4.3}$$

The matrices $P^*\big(\lambda_1(\varepsilon)\big)$, $M_1(\varepsilon)$, and M_2 are all clearly bounded as $\varepsilon \downarrow 0$, and consequently the reasoning used in Proposition 4.4 below can be used to show that the first three terms in (4.3) all go to 0. For the fourth term, we make use of the fact that, for $\varepsilon \neq 0$, $\varepsilon M_3(\varepsilon)$ is a constant matrix:

$$\begin{aligned} &\frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)} \big(\lambda_2^k(\varepsilon) - \lambda_3^k(\varepsilon)\big)M_3(\varepsilon) \\ &\qquad = \frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)} \big(\lambda_2(\varepsilon) - \lambda_3(\varepsilon)\big)\sum_{j=1}^{k} \lambda_2^{k-j}(\varepsilon)\lambda_3^{j-1}(\varepsilon)M_3(\varepsilon) \end{aligned}$$

$$= \frac{1}{N(\varepsilon)} \frac{\lambda_2(\varepsilon) - \lambda_3(\varepsilon)}{\varepsilon} \sum_{k=1}^{N(\varepsilon)} \sum_{j=1}^{k} \lambda_2^{k-j}(\varepsilon) \lambda_3^{j-1}(\varepsilon) \left[\varepsilon M_3(\varepsilon)\right]$$

$$= \frac{2\varepsilon^2 - 2\varepsilon}{N(\varepsilon)} \sum_{j=1}^{N(\varepsilon)} \lambda_3^{j-1}(\varepsilon) \sum_{k=1}^{N(\varepsilon)-j+1} \lambda_2^{k-1}(\varepsilon) \left[\varepsilon M_3(\varepsilon)\right]$$

$$= \frac{2\varepsilon^2 - 2\varepsilon}{N(\varepsilon)} \sum_{j=1}^{N(\varepsilon)} \lambda_3^{j-1}(\varepsilon) \frac{1 - \lambda_2^{N(\varepsilon)-j+1}(\varepsilon)}{1 - \lambda_2(\varepsilon)} \left[\varepsilon M_3(\varepsilon)\right]$$

$$= \frac{1}{N(\varepsilon)} \frac{2\varepsilon^2 - 2\varepsilon}{2 - 2\varepsilon} \sum_{j=1}^{N(\varepsilon)} \lambda_3^{j-1}(\varepsilon) \left(1 - \lambda_2^{N(\varepsilon)-j+1}(\varepsilon)\right) \left[\varepsilon M_3(\varepsilon)\right]$$

$$= -\frac{\varepsilon}{N(\varepsilon)} \sum_{j=1}^{N(\varepsilon)} \lambda_3^{j-1}(\varepsilon) \left(1 - \lambda_2^{N(\varepsilon)-j+1}(\varepsilon)\right) \left[\varepsilon M_3(\varepsilon)\right].$$

Hence

$$\left\| \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} \left(\lambda_2^k(\varepsilon) - \lambda_3^k(\varepsilon)\right) M_3(\varepsilon) \right\|$$

$$= \left\| -\frac{\varepsilon}{N(\varepsilon)} \sum_{j=1}^{N(\varepsilon)} \lambda_3^{j-1}(\varepsilon) \left(1 - \lambda_2^{N(\varepsilon)-j+1}(\varepsilon)\right) \left[\varepsilon M_3(\varepsilon)\right] \right\|$$

$$= \frac{\varepsilon}{N(\varepsilon)} \left| \sum_{j=1}^{N(\varepsilon)} \lambda_3^{j-1}(\varepsilon) \left(1 - \lambda_2^{N(\varepsilon)-j+1}(\varepsilon)\right) \right| \left\| \varepsilon M_3(\varepsilon) \right\|$$

$$\leq \frac{\varepsilon}{N(\varepsilon)} \sum_{j=1}^{N(\varepsilon)} \left| \lambda_3^{j-1}(\varepsilon) \left(1 - \lambda_2^{N(\varepsilon)-j+1}(\varepsilon)\right) \right| \left\| \varepsilon M_3(\varepsilon) \right\|$$

$$\leq \frac{\varepsilon}{N(\varepsilon)} \sum_{j=1}^{N(\varepsilon)} (1 \cdot 2) \left\| \varepsilon M_3(\varepsilon) \right\|$$

$$= 2\varepsilon \left\| \varepsilon M_3(\varepsilon) \right\|.$$

Since $2\varepsilon \left\| \varepsilon M_3(\varepsilon) \right\| \to 0$ as $\varepsilon \downarrow 0$, it follows that the fourth term from (4.3) goes to 0 along with the other three, implying the result.

4.3 Further Decomposition of the Limit

Suppose that λ is an eigenvalue of P_0 that lies on the unit circle. In this section we describe how the total projection $P^*(\lambda, \varepsilon)$, which appears in the initial decomposition of the hybrid Cesaro limit from Section 4.1, can be further broken down into more readily-analyzed pieces. We first express $P^*(\lambda, \varepsilon)$ as

$$P^*(\lambda, \varepsilon) = P^*_{k(\lambda)}(\lambda, \varepsilon) + \sum_{i=1}^{k(\lambda)} (P^*_{i-1}(\lambda, \varepsilon) - P^*_i(\lambda, \varepsilon)) \tag{4.4}$$

for certain analytic projection matrices $P^*_i(\lambda, \varepsilon)$. (Here $k(\lambda)$ is a nonnegative integer that will emerge from the process.) Subsequently, for each $i \leq k(\lambda)$ we will decompose $P^*_{i-1}(\lambda, \varepsilon) - P^*_i(\lambda, \varepsilon)$.

To this end, we begin by introducing notation that allows us to describe the decomposition process. Let $m_0(\lambda)$ denote the multiplicity of the eigenvalue 0 in $P_0 - \lambda I$, so that $m_0(\lambda)$ is the same as the multiplicity of λ in P_0. Let $m'(\lambda)$ denote the multiplicity of 0 in $T_0(\lambda, \varepsilon) = P(\varepsilon) - \lambda I = (P_0 - \lambda I) + A(\varepsilon)$. That is, if a perturbed eigenvalue of $T_0(\lambda, \varepsilon)$ is identically equal to 0, then $m'(\lambda)$ is its multiplicity; otherwise, $m'(\lambda) = 0$. Note that $m'(\lambda)$ is the same as the multiplicity of λ in $P(\varepsilon)$. In each of Example 4.1 and Example 4.2, then, $m'(-1) = 1$. Let $P^*_0(\lambda, \varepsilon)$ be the total projection matrix for the 0-group eigenvalues in $T_0(\lambda, \varepsilon)$; then $T_0(\lambda, \varepsilon)$ is identical to $P^*(\lambda, \varepsilon)$, the total projection matrix for the λ-group eigenvalues in $P(\varepsilon)$. Also, let $M_0(\lambda, \varepsilon) = M(\lambda, \varepsilon)$, the $m_0(\lambda)$-dimensional perturbed eigenspace for the λ-group in $P(\varepsilon)$. Notice that $M_0(\lambda, \varepsilon)$ is the same as the perturbed eigenspace for the 0-group in $T_0(\lambda, \varepsilon)$. Finally, let $M_{-1}(\lambda, \varepsilon) = \mathbb{C}^n$, so that $T_0(\lambda, \varepsilon)$ acts on $M_{-1}(\lambda, \varepsilon)$ in the sense introduced in Section 3.2, and observe that $m_0(\lambda) \geq m'(\lambda)$. We now proceed inductively as follows.

Given $T_j(\lambda, \varepsilon)$, which acts on $M_{j-1}(\lambda, \varepsilon)$, and the corresponding analytic projection $P^*_j(\lambda, \varepsilon)$ onto the $m_j(\lambda)$-dimensional subspace $M_j(\lambda, \varepsilon)$, one of the following three conditions must hold:

(1) 0 is not a semisimple eigenvalue of $T_j(\lambda, 0)$ in $M_{j-1}(\lambda, 0)$;

(2) $m_j(\lambda) = m'(\lambda)$; or

(3) 0 is a semisimple eigenvalue of $T_j(\lambda, 0)$ in $M_{j-1}(\lambda, 0)$.and $m_j(\lambda) > m'(\lambda)$.

In the first two cases we terminate the decomposition process: we let $k(\lambda) = j$, and take the decomposition of $P^*(\lambda, \varepsilon)$ as in (4.4). In the third case, we apply the reduction process as described in Section 3.2: we let

$$T_{j+1}(\lambda, \varepsilon) = \frac{1}{\varepsilon} T_j(\lambda, \varepsilon) P^*_j(\lambda, \varepsilon),$$

so that $T_{j+1}(\lambda,\varepsilon)$ acts on the subspace $M_j(\lambda,\varepsilon)$. We now define $M_{j+1}(\lambda,\varepsilon)$ to be the intersection of $M_j(\lambda,\varepsilon)$ with the eigenspace for the 0-group of eigenvalues in $T_{j+1}(\lambda,\varepsilon)$, and we define $P^*_{j+1}(\lambda,\varepsilon)$ to be the analytic projection onto this subspace. Letting $m_{j+1}(\lambda)$ denote the dimension of this subspace, we see that $m_j(\lambda) \geq m_{j+1}(\lambda) \geq m'(\lambda)$.

This process is guaranteed to halt after some finite number of steps. For suppose that condition (1) above never occurs, and let $\lambda(\varepsilon) = \lambda + \sum_{l=1}^{\infty} c_l \varepsilon^{l/p}$ be the Puiseux series for a perturbed eigenvalue in the λ-group of $P(\varepsilon)$ that is not identically equal to λ. Also, let l_0 be the smallest value of l for which $c_l \neq 0$. Then by what was noted in Section 3.2 about how the reduction process affects perturbed eigenvalues, for each nonnegative integer $j < l_0/p$ we have that

$$\frac{1}{\varepsilon^j}(\lambda(\varepsilon)-\lambda) = \sum_{l=1}^{\infty} c_l \varepsilon^{(l/p)-j} = \sum_{l=l_0}^{\infty} c_l \varepsilon^{(l/p)-j}$$

is the Puiseux series for a perturbed eigenvalue in the 0-group of $T_j(\lambda,\varepsilon)$. In particular, this is true for the largest integer j' less than l_0/p. Letting $l_1 = p(j'+1)$, we see that

$$\frac{1}{\varepsilon}\sum_{l=l_0}^{\infty} c_l \varepsilon^{(l/p)-j'} = \sum_{l=l_0}^{l_1-1} c_l \varepsilon^{(l/p)-(j'+1)} + c_{l_1} + \sum_{l=l_1+1}^{\infty} c_l \varepsilon^{(l/p)-(j'+1)} \tag{4.5}$$

is a perturbed eigenvalue of $T_{j'+1}(\lambda,\varepsilon)$, and hence must approach some eigenvalue of $T_{j'+1}(\lambda,0)$ as $\varepsilon \downarrow 0$. But this necessitates that $l_1 = l_0$, for otherwise the first sum on the right-hand side of the above identity would diverge as $\varepsilon \downarrow 0$. (The second sum approaches 0 as a result of the way we defined l_1.) Thus the perturbed eigenvalue in (4.5) is in the c_{l_0}-group of $T_{j'+1}(\lambda,\varepsilon)$, whence $m_{j'+1}(\lambda) < m_{j'}(\lambda)$. From all of this we see that the $m_j(\lambda)$ must eventually decrease to $m'(\lambda)$, as every perturbed λ-group eigenvalue not identically equal to λ is eventually "split off" in the reduction process.

Before proceeding with the second step of the decomposition, we introduce a term to distinguish between the two conditions under which the process laid out above terminates. If condition (2) is eventually satisfied, we say that λ is *completely reducible* for $P(\varepsilon)$. In [4], the authors show that 1 is completely reducible for any perturbed stochastic matrix $P(\varepsilon)$.

Continuing with the decomposition, suppose that $1 \leq i \leq k(\lambda)$ and that $m_i(\lambda) < m_{i-1}(\lambda)$. Then $M_i(\lambda,\varepsilon)$ is a proper subspace of $M_{i-1}(\lambda,\varepsilon)$, so the analytic projections $P^*_i(\lambda,\varepsilon)$ and $P^*_{i-1}(\lambda,\varepsilon)$ onto these subspaces are not equal. (If instead $m_i(\lambda) = m_{i-1}(\lambda)$, then the subspaces and hence the analytic projections are identical.) In this case, the unperturbed reduced matrix $T_i(\lambda, 0)$ must have a nonempty collection of nonzero eigenvalues, call them $\mu_{i,1}(\lambda)$, $\mu_{i,2}(\lambda)$, $\ldots$, $\mu_{i,n_i(\lambda)}(\lambda)$; these $\mu_{i,j}(\lambda)$ are just the eigenvalues that "split off" from 0 after the $(i+1)$st application of the reduction process.) For $1 \le j \le n_i(\lambda)$, we let $P^*_{i,j}(\lambda, \varepsilon)$ be the total projection for the $\mu_{i,j}(\lambda)$-group eigenvalues in $T_i(\lambda, \varepsilon)$. Then

$$P^*_{i-1}(\lambda, \varepsilon) = P^*_i(\lambda, \varepsilon) + \sum_{j=1}^{n_i(\lambda)} P^*_{i,j}(\lambda, \varepsilon).$$

With this secondary decomposition, we arrive at our desired decomposition of $P^*(\lambda, \varepsilon)$:

$$\begin{aligned} P^*(\lambda, \varepsilon) &= P^*_{k(\lambda)}(\lambda, \varepsilon) + \sum_{i=1}^{k(\lambda)} (P^*_{i-1}(\lambda, \varepsilon) - P^*_i(\lambda, \varepsilon)) \\ &= P^*_{k(\lambda)}(\lambda, \varepsilon) + \sum_{i=1}^{k(\lambda)} \sum_{j=1}^{n_i(\lambda)} P^*_{i,j}(\lambda, \varepsilon). \end{aligned} \tag{4.6}$$

As mentioned earlier, we make use of this decomposition in evaluating smaller pieces of the overall Cesaro limit expression. Before continuing, though, note that if $0 \le i_1 < i_2 \le k(\lambda)$, then $P^*_{i_2}(\lambda, \varepsilon)$ is a sub-projection of $P^*_{i_1}(\lambda, \varepsilon)$. Similarly, for such i_1 and i_2, if $1 \le j \le n_{i_2}(\lambda)$, then $P^*_{i_2,j}(\lambda, \varepsilon)$ is a sub-projection of $P^*_{i_1}(\lambda, \varepsilon)$. Hence, by the inductive definition of the $T_i(\lambda, \varepsilon)$, we see that

$$\begin{aligned} T_i(\lambda, \varepsilon) &= \frac{1}{\varepsilon} T_{i-1}(\lambda, \varepsilon) P^*_{i-1}(\lambda, \varepsilon) = \frac{1}{\varepsilon}\left[\frac{1}{\varepsilon} T_{i-2}(\lambda, \varepsilon) P^*_{i-2}(\lambda, \varepsilon)\right] P^*_{i-1}(\lambda, \varepsilon) \\ &= \frac{1}{\varepsilon^2} T_{i-2}(\lambda, \varepsilon) P^*_{i-2}(\lambda, \varepsilon) P^*_{i-1}(\lambda, \varepsilon) = \frac{1}{\varepsilon^2} T_{i-2}(\lambda, \varepsilon) P^*_{i-1}(\lambda, \varepsilon) \\ &= \cdots = \frac{1}{\varepsilon^j} T_{i-j}(\lambda, \varepsilon) P^*_{i-1}(\lambda, \varepsilon) \\ &= \cdots = \frac{1}{\varepsilon^i} T_0(\lambda, \varepsilon) P^*_{i-1}(\lambda, \varepsilon) = \frac{1}{\varepsilon^i} (P(\varepsilon) - \lambda I) P^*_{i-1}(\lambda, \varepsilon). \end{aligned}$$

4.4 Previous Results

We begin by introducing a piece of notation. If λ is a unit-circle eigenvalue of P_0, $1 \le i \le k(\lambda)$, and $1 \le j \le n_i(\lambda)$, we define

$$D_{i,j}(\lambda, \varepsilon) = (T_i(\lambda, \varepsilon) - \mu_{i,j}(\lambda) I) P^*_{i,j}(\lambda, \varepsilon).$$

We refer to these matrices as generalized analytic idempotents. Note that each $D_{i,j}(\lambda, 0)$ is a nilpotent matrix; we denote its index of nilpotence by $n_{i,j}(\lambda)$.

The three propositions below, which are restatements of results in [4], are central to the overall characterization of the hybrid Cesaro limit there. All three concern the projection matrices obtained as above by decomposing $P^*(1, \varepsilon)$.

Proposition 4.1. *If $1 \le i \le k(1)$ and $N(\varepsilon)\varepsilon^i \to \infty$ as $\varepsilon \downarrow 0$, then for $1 \le j \le n_i(1)$,*

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*_{i,j}(1, \varepsilon) = 0.$$

Proposition 4.2. *If $1 \le i \le k(1)$ and $N(\varepsilon)\varepsilon^i \to 0$ as $\varepsilon \downarrow 0$, then for $1 \le j \le n_i(1)$,*

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*_{i,j}(1, \varepsilon) = P^*_{i,j}(1, 0).$$

Proposition 4.3. *If $1 \le i \le k(1)$ and $N(\varepsilon)\varepsilon^i \to L$ as $\varepsilon \downarrow 0$, where $0 < L < \infty$, then for $1 \le j \le n_i(1)$,*

$$\begin{aligned}
&\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*_{i,j}(1, \varepsilon) \\
&\quad = \frac{1 - e^{L\mu_{i,j}(1)}}{-L\mu_{i,j}(1)} P^*_{i,j}(1, 0) \\
&\qquad + \frac{1}{-L\mu_{i,j}(1)} \sum_{l=1}^{n_{i,j}(1)-1} \left(1 - e^{L\mu_{i,j}(1)} \sum_{k=0}^{l} \frac{\big(-L\mu_{i,j}(1)\big)^k}{k!} \right) \\
&\qquad\quad \times \frac{D^l_{i,j}(1, 0)}{\big(-\mu_{i,j}(1)\big)^l}.
\end{aligned}$$

Now, since 1 is completely reducible for $P(\varepsilon)$, $P^*_{k(1)}(1, \varepsilon)$ projects onto the eigenspace for 1 in $P(\varepsilon)$, so $P(\varepsilon) P^*_{k(1)}(1, \varepsilon) = P^*_{k(1)}(1, \varepsilon)$. Consequently,

$$\begin{aligned}
\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*_{k(1)}(1, \varepsilon) &= \lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^*_{k(1)}(1, \varepsilon) \\
&= \lim_{\varepsilon \downarrow 0} P^*_{k(1)}(1, \varepsilon) = P^*_{k(1)}(1, 0).
\end{aligned}$$

The following three major results (essentially Theorems 1, 2, and 3 in [4], p. 237) are a consequence of the above propositions and the observation just made.

Theorem 4.1. *Suppose that* $0 \le i \le k(1) - 1$ *and* $N(\varepsilon) \uparrow \infty$ *with* $N(\varepsilon)\varepsilon^i \to \infty$ *but* $N(\varepsilon)\varepsilon^{i+1} \to 0$ *as* $\varepsilon \downarrow 0$. *Then*

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(1, \varepsilon) = P_i^*(1, 0).$$

Theorem 4.2. *If* $N(\varepsilon)\varepsilon^{k(1)} \to \infty$ *as* $\varepsilon \downarrow 0$, *then*

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(1, \varepsilon) = P_k^*(1, 0).$$

Theorem 4.3. *Suppose* $1 \le i \le k(1)$ *and* $N(\varepsilon)\varepsilon^i \to L$ *as* $\varepsilon \downarrow 0$, *where* $0 < L < \infty$. *Then*

$$\begin{aligned}
&\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(1, \varepsilon) \\
&\quad = P_i^*(1, 0) + \sum_{j=1}^{n_i(1)} \frac{1 - e^{L\mu_{i,j}(1)}}{-L\mu_{i,j}(1)} P_{i,j}^*(1, 0) \\
&\qquad + \sum_{j=1}^{n_i(1)} \frac{1}{-L\mu_{i,j}(1)} \sum_{l=1}^{n_{i,j}(1)-1} \left(1 - e^{L\mu_{i,j}(1)} \sum_{k=0}^{l} \frac{\big(-L\mu_{i,j}(1)\big)^k}{k!} \right) \frac{D_{i,j}^l(1, 0)}{\big(-\mu_{i,j}(1)\big)^l}
\end{aligned}$$

The final result which we present from [4] concerns eigenvalues λ of P_0 satisfying $|\lambda| < 1$. It follows from Proposition 1, p. 235 there.

Theorem 4.4. *Suppose that* λ *is an eigenvalue of* P_0 *such that* $|\lambda| < 1$. *Then*

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(\lambda, \varepsilon) = 0.$$

4.5 New Results

The previous determinations of the value of

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(\lambda, \varepsilon), \tag{4.7}$$

when either $\lambda = 1$ or $|\lambda| < 1$, do not change when P_0 is permitted to have eigenvalues on the unit circle other than 1. Therefore to fully characterize the overall hybrid Cesaro limit, it suffices to concentrate on the limit in (4.7) when λ is an eigenvalue on the unit circle other than 1; that is, we wish to determine whether this limit exists, and if it does we wish to determine a general expression for it. One result I obtained, from my work over the summer (see [8]), is as follows.

Proposition 4.4. *Suppose that λ is an eigenvalue of P_0 satisfying $|\lambda| = 1$, $\lambda \neq 1$, and let $\lambda_1(\varepsilon), \ldots, \lambda_m(\varepsilon)$ be the perturbed λ-group eigenvalues, with associated eigenprojections $P^*\big(\lambda_1(\varepsilon)\big), \ldots, P^*\big(\lambda_m(\varepsilon)\big)$. If each $\lambda_i(\varepsilon)$ is a semisimple eigenvalue of $P(\varepsilon)$ and each $P^*\big(\lambda_i(\varepsilon)\big)$ is bounded as $\varepsilon \downarrow 0$, then*

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(\lambda, \varepsilon) = 0.$$

Proof. Since the $\lambda_i(\varepsilon)$ are semisimple,

$$\begin{aligned}
\frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(\lambda, \varepsilon) &= \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) \sum_{i=1}^{m} P^*\big(\lambda_i(\varepsilon)\big) \\
&= \sum_{i=1}^{m} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*\big(\lambda_i(\varepsilon)\big) \\
&= \sum_{i=1}^{m} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} \lambda_i^k(\varepsilon) P^*\big(\lambda_i(\varepsilon)\big) \\
&= \sum_{i=1}^{m} \frac{\lambda_i(\varepsilon)}{N(\varepsilon)} \frac{1 - \lambda_i^{N(\varepsilon)}(\varepsilon)}{1 - \lambda_i(\varepsilon)} P^*\big(\lambda_i(\varepsilon)\big).
\end{aligned}$$

By hypothesis, there exist positive constants $\varepsilon_0, M_1, \ldots, M_m$ such that for all $\varepsilon \in (0, \varepsilon_0)$, $P(\varepsilon)$ is stochastic and $\big\|P^*\big(\lambda_i(\varepsilon)\big)\big\| \leq M_i$, $i = 1, \ldots, m$. Hence, for these ε, $|\lambda_i(\varepsilon)| \leq 1$, $i = 1, \ldots, m$. It follows that

$$\begin{aligned}
\left\| \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(\lambda, \varepsilon) \right\| &\leq \sum_{i=1}^{m} \left\| \frac{\lambda_i(\varepsilon)}{N(\varepsilon)} \frac{1 - \lambda_i^{N(\varepsilon)}(\varepsilon)}{1 - \lambda_i(\varepsilon)} P^*\big(\lambda_i(\varepsilon)\big) \right\| \\
&= \sum_{i=1}^{m} \frac{|\lambda_i(\varepsilon)|}{N(\varepsilon)} \frac{|1 - \lambda_i^{N(\varepsilon)}(\varepsilon)|}{|1 - \lambda_i(\varepsilon)|} \big\|P^*\big(\lambda_i(\varepsilon)\big)\big\| \\
&\leq \sum_{i=1}^{m} \frac{1}{N(\varepsilon)} \frac{1 + |\lambda_i(\varepsilon)|^{N(\varepsilon)}}{|1 - \lambda_i(\varepsilon)|} M_i
\end{aligned}$$

$$\leq \frac{1}{N(\varepsilon)} \sum_{i=1}^{m} \frac{2M_i}{|1-\lambda_i(\varepsilon)|},$$

again for $\varepsilon \in (0, \varepsilon_0)$. As

$$\sum_{i=1}^{m} \frac{2M_i}{|1-\lambda_i(\varepsilon)|} \to \frac{2}{|1-\lambda|} \sum_{i=1}^{m} M_i$$

but $N(\varepsilon) \to \infty$ as $\varepsilon \downarrow 0$,

$$\left\| \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(\lambda, \varepsilon) \right\| \to 0,$$

and the result follows. □

Note that each $P^*\big(\lambda_i(\varepsilon)\big)$ above need not be an analytic perturbation of any matrix.

We now come to the main results. Throughout, we assume that $\lambda \neq 1$ is a unit-circle eigenvalue of P_0. If, in the process of reducing $P(\varepsilon)$ for λ, we have that $\alpha_{i,j}(\lambda, \varepsilon)$ is a perturbed eigenvalue for $T_i(\lambda, \varepsilon)$ in the $\mu_{i,j}(\lambda)$-group ($1 \leq i \leq k(\lambda)$ and $1 \leq j \leq n_i(\lambda)$), then $\beta_{i,j}(\lambda, \varepsilon) = \lambda + \varepsilon^i \alpha_{i,j}(\lambda, \varepsilon) = \lambda + \varepsilon^i \mu_{i,j}(\lambda) + \cdots$ is the corresponding perturbed eigenvalue in $P(\varepsilon)$. Since $P(\varepsilon)$ is stochastic for all sufficiently small positive ε, say $0 < \varepsilon < \varepsilon_0$, it must be the case that $\beta_{i,j}(\lambda, \varepsilon)$ lies within Θ_n for such ε. As the term $\varepsilon^i \mu_{i,j}(\lambda)$ in the Puiseux series for $\beta_{i,j}(\lambda, \varepsilon)$ dominates all the other non-constant terms as $\varepsilon \downarrow 0$, certainly $\operatorname{Re}[\overline{\lambda} \mu_{i,j}(\lambda)] \leq 0$: otherwise, $\beta_{i,j}(\lambda, \varepsilon)$ would lie strictly outside the unit disk for all sufficiently small positive ε, whereas Θ_n is entirely contained within the unit disk.

In fact, however, the results from Section 2.3 imply that strict inequality holds above. For suppose to the contrary that $\operatorname{Re}[\overline{\lambda} \mu_{i,j}(\lambda)] = 0$. Then the curve traced out by $\beta_{i,j}(\lambda, \varepsilon)$ in the complex plane for $0 \leq \varepsilon < \varepsilon_0$ is tangent to the unit circle at λ. But the boundary arcs of Θ_n incident to λ make nonzero angles with the unit circle at λ by Theorems 2.3 and 2.4, so it is necessarily the case that $\beta_{i,j}(\lambda, \varepsilon)$ falls outside of Θ_n for all sufficiently small positive ε. As this cannot happen, the claimed strict inequality must hold. Now define $\nu_{i,j}(\lambda, \varepsilon) = \lambda + \varepsilon^i \mu_{i,j}(\lambda)$, and write $\overline{\lambda} \mu_{i,j}(\lambda) = -a + bi$, so that $a > 0$. Then for positive ε so small that

$$\left(\frac{3}{4} a^2 + b^2 \right) \varepsilon^i \leq a,$$

we obtain

$$\begin{aligned}
\left|\nu_{i,j}(\lambda,\varepsilon)\right| &= \left|\nu_{i,j}(\lambda,\varepsilon)\right|\left|\overline{\lambda}\right| = \left|\left(\lambda+\varepsilon^{i}\mu_{i,j}(\lambda)\right)\overline{\lambda}\right| \\
&= \left|1+\varepsilon^{i}\overline{\lambda}\mu_{i,j}(\lambda)\right| \\
&\leq 1+\frac{1}{2}\varepsilon^{i}\mathrm{Re}\,[\overline{\lambda}\mu_{i,j}(\lambda)] < 1.
\end{aligned}$$

Additionally, by what was noted at the end of Section 4.3 we have the identity

$$\begin{aligned}
\left[P(\varepsilon)-\nu_{i,j}(\lambda,\varepsilon)I\right]P^{*}_{i,j}(\lambda,\varepsilon) &= \left[P(\varepsilon)-\lambda I-\varepsilon^{i}\mu_{i,j}(\lambda)I\right]P^{*}_{i,j}(\lambda,\varepsilon) \\
&= \left[(P(\varepsilon)-\lambda I)P^{*}_{i-1}(\lambda,\varepsilon)-\varepsilon^{i}\mu_{i,j}(\lambda)I\right]P^{*}_{i,j}(\lambda,\varepsilon) \\
&= \varepsilon^{i}\left[\frac{1}{\varepsilon^{i}}(P(\varepsilon)-\lambda I)P^{*}_{i-1}(\lambda,\varepsilon)-\mu_{i,j}(\lambda)I\right]P^{*}_{i,j}(\lambda,\varepsilon) \\
&= \varepsilon^{i}\left[T_{i}(\lambda,\varepsilon)-\mu_{i,j}(\lambda)I\right]P^{*}_{i,j}(\lambda,\varepsilon) \\
&= \varepsilon^{i}D_{i,j}(\lambda,\varepsilon).
\end{aligned}$$

As a consequence of this and the fact that

$$P^{*}_{i,j}(\lambda,\varepsilon)D_{i,j}(\lambda,\varepsilon) = D_{i,j}(\lambda,\varepsilon)P^{*}_{i,j}(\lambda,\varepsilon) = D_{i,j}(\lambda,\varepsilon),$$

we obtain the following lemma (this is virtually identical to Lemma 2 in [4]).

Lemma 4.1.

$$\begin{aligned}
\frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)}P^{k}(\varepsilon)P^{*}_{i,j}(\lambda,\varepsilon) &= \frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)}\nu^{k}_{i,j}(\lambda,\varepsilon)P^{*}_{i,j}(\lambda,\varepsilon) \\
&\quad + \frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)}\sum_{l=1}^{k}\binom{k}{l}\nu^{k-l}_{i,j}[\varepsilon^{i}D_{i,j}(\lambda,\varepsilon)]^{l}.
\end{aligned}$$

Proof. Since $P^{*}_{i,j}(\lambda,\varepsilon)$ is a projection matrix and commutes with $P(\varepsilon)$, we have that

$$P^{k}(\varepsilon)P^{*}_{i,j}(\lambda,\varepsilon) = P^{k}(\varepsilon)[P^{*}_{i,j}(\lambda,\varepsilon)]^{k} = [P(\varepsilon)P^{*}_{i,j}(\lambda,\varepsilon)]^{k}.$$

Using what was noted before the lemma, we see that

$$\begin{aligned}
P^{k}(\varepsilon)P^{*}_{i,j}(\lambda,\varepsilon) &= [\nu_{i,j}(\lambda,\varepsilon)P^{*}_{i,j}(\lambda,\varepsilon)+\varepsilon^{i}D_{i,j}(\lambda,\varepsilon)]^{k} \\
&= \sum_{l=0}^{k}\binom{k}{l}[\nu_{i,j}(\lambda,\varepsilon)P^{*}_{i,j}(\lambda,\varepsilon)]^{k-l}[\varepsilon^{i}D_{i,j}(\lambda,\varepsilon)]^{l}
\end{aligned}$$

$$
\begin{aligned}
&= [\nu_{i,j}(\lambda,\varepsilon)P^*_{i,j}(\lambda,\varepsilon)]^k \\
&\qquad + \sum_{l=1}^{k}\binom{k}{l}[\nu_{i,j}(\lambda,\varepsilon)P^*_{i,j}(\lambda,\varepsilon)]^{k-l}[\varepsilon^i D_{i,j}(\lambda,\varepsilon)]^l \\
&= \nu^k_{i,j}(\lambda,\varepsilon)P^*_{i,j}(\lambda,\varepsilon) + \sum_{l=1}^{k}\binom{k}{l}\nu^{k-l}_{i,j}(\lambda,\varepsilon)[\varepsilon^i D_{i,j}(\lambda,\varepsilon)]^l.
\end{aligned}
$$

The result now follows by summing over k and dividing by $N(\varepsilon)$. □

So for each $P^*_{i,j}(\lambda,\varepsilon)$, this lemma further breaks down the expression of interest. We deal with the first piece immediately, and then return to the second, obtaining several useful estimates, before proving the main propositions.

Lemma 4.2.

$$
\lim_{\varepsilon\downarrow 0}\frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)}\nu^k_{i,j}(\lambda,\varepsilon)P^*_{i,j}(\lambda,\varepsilon) = 0.
$$

Proof. Using the expression for a finite geometric series, we estimate the norm of the limit expression as follows:

$$
\begin{aligned}
&\left\|\frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)}\nu^k_{i,j}(\lambda,\varepsilon)P^*_{i,j}(\lambda,\varepsilon)\right\| \\
&= \left\|\frac{1}{N(\varepsilon)}\,\frac{\nu_{i,j}(\lambda,\varepsilon)\left[1-\nu^{N(\varepsilon)}_{i,j}(\lambda,\varepsilon)\right]}{1-\nu_{i,j}(\lambda,\varepsilon)}P^*_{i,j}(\lambda,\varepsilon)\right\| \\
&\le \frac{1}{N(\varepsilon)}\left|\frac{\nu_{i,j}(\lambda,\varepsilon)\left[1-\nu^{N(\varepsilon)}_{i,j}(\lambda,\varepsilon)\right]}{1-\nu_{i,j}(\lambda,\varepsilon)}\right|\|P^*_{i,j}(\lambda,\varepsilon)\| \\
&\le \frac{2}{N(\varepsilon)|1-\nu_{i,j}(\lambda,\varepsilon)|}\|P^*_{i,j}(\lambda,\varepsilon)\|.
\end{aligned}
$$

But $\|P^*_{i,j}(\lambda,\varepsilon)\| \to \|P^*_{i,j}(\lambda,0)\|$ and $|1-\nu_{i,j}(\lambda,\varepsilon)| \to |1-\lambda| \neq 0$ as $\varepsilon\downarrow 0$, so the fact that $N(\varepsilon)\uparrow\infty$ implies that the final expression above approaches 0 as $\varepsilon\downarrow 0$. Hence the first expression also approaches 0, and the lemma follows. □

Returning to the second term from Lemma 4.1, the next three lemmas allow us to estimate this term. They are taken almost unchanged from [4], pp. 238–239.

Lemma 4.3.

$$\frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} \sum_{l=1}^{k} \binom{k}{l} \nu_{i,j}^{k-l}(\lambda,\varepsilon) \left[\varepsilon^i D_{i,j}(\lambda,\varepsilon)\right]^l$$
$$= \frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \left[\sum_{k=l}^{N(\varepsilon)} \binom{k}{l} \nu_{i,j}^{k-l}(\lambda,\varepsilon) \right] \left[\varepsilon^i D_{i,j}(\lambda,\varepsilon)\right],$$

and

$$\sum_{k=l}^{N(\varepsilon)} \binom{k}{l} \nu_{i,j}^{k-l}(\lambda,\varepsilon) = \frac{1 - \sum_{k=0}^{l} \binom{N(\varepsilon)+1}{k} \nu_{i,j}^{N(\varepsilon)+1-k}(\lambda,\varepsilon) \left[1-\nu_{i,j}(\lambda,\varepsilon)\right]^k}{\left[1-\nu_{i,j}(\lambda,\varepsilon)\right]^l}.$$

Lemma 4.4.

$$\left| \sum_{k=l}^{N(\varepsilon)} \binom{k}{l} \nu_{i,j}^{k-l}(\varepsilon) \right| = \left| \sum_{k=0}^{N(\varepsilon)-l} \binom{k+l}{l} \nu_{i,j}^{k}(\varepsilon) \right| \le \sum_{k=0}^{N(\varepsilon)-l} \binom{k+l}{l} \left|\nu_{i,j}(\varepsilon)\right|^k.$$

Also, we have the two estimates

$$\sum_{k=0}^{N(\varepsilon)-l} \binom{k+l}{l} \left|\nu_{i,j}(\varepsilon)\right|^k \le \sum_{k=0}^{\infty} \binom{k+l}{l} \left|\nu_{i,j}(\varepsilon)\right|^k = \frac{1}{\left[1-\left|\nu_{i,j}(\varepsilon)\right|\right]^{l+1}}$$

and

$$\sum_{k=0}^{N(\varepsilon)-l} \binom{k+l}{l} \left|\nu_{i,j}(\varepsilon)\right|^k \le \sum_{k=0}^{N(\varepsilon)-l} \binom{k+l}{l} = \sum_{k=0}^{N(\varepsilon)-l} \frac{(l+k)\cdots(1+k)}{l!}$$
$$\le \frac{N^{l+1}(\varepsilon)}{l!},$$

where the latter is valid for $l \ge 1$.

Lemma 4.5. *If $1 \le i \le k(\lambda)$ and $1 \le j \le n_i(\lambda)$, then there are positive constants $C_{i,j}(\lambda)$, $K_{i,j}(\lambda)$, and $\varepsilon_{i,j}(\lambda)$ such that*

$$\left\|D_{i,j}^l(\lambda,\varepsilon)\right\| \le C_{i,j}(\lambda)\left(K_{i,j}(\lambda)\varepsilon\right)^{l/n_{i,j}^2(\lambda)}$$

whenever $0 < \varepsilon < \varepsilon_{i,j}(\lambda)$ and $l \ge n_{i,j}(\lambda)$. Also, for sufficiently small positive ε we have $\left\|D_{i,j}(\lambda,\varepsilon)\right\| \le \left\|D_{i,j}(\lambda,0)\right\| + 1$.

We now present the main propositions.

Proposition 4.5. *If* $1 \le i \le k(\lambda)$ *and* $N(\varepsilon)\varepsilon^i \to \infty$ *as* $\varepsilon \downarrow 0$*, then for* $1 \le j \le n_i(\lambda)$*,*

$$\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*_{i,j}(\lambda,\varepsilon) = 0.$$

Proof. Suppose that $1 \le i \le k(\lambda)$ and $1 \le j \le n_i(\lambda)$ with $N(\varepsilon)\varepsilon^i \to \infty$ as $\varepsilon \downarrow 0$. By the results in Lemmas 4.1, 4.2, and 4.3, it suffices to show that

$$\frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \left[\sum_{k=l}^{N(\varepsilon)} \binom{k}{l} \nu_{i,j}^{k-l}(\lambda,\varepsilon) \right] \left[\varepsilon^i D_{i,j}(\lambda,\varepsilon)\right]^l \to 0$$

as $\varepsilon \to 0$. We will estimate the norm of this expression using the first estimate from Lemma 4.4, the estimates from Lemma 4.5, and the fact (noted at the beginning of this chapter) that for all sufficiently small positive ε,

$$\left|\nu_{i,j}(\lambda,\varepsilon)\right| \le 1 + \frac{1}{2}\varepsilon^i \mathrm{Re}\,[\bar{\lambda}\mu_{i,j}(\lambda)].$$

Thus for small enough ε,

$$\begin{aligned}
&\left\| \frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \left[\sum_{k=l}^{N(\varepsilon)} \binom{k}{l} \nu_{i,j}^{k-l}(\lambda,\varepsilon) \right] \left[\varepsilon^i D_{i,j}(\lambda,\varepsilon)\right]^l \right\| \\
&\quad \le \frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \left| \sum_{k=l}^{N(\varepsilon)} \binom{k}{l} \nu_{i,j}^{k-l}(\lambda,\varepsilon) \right| \varepsilon^{li} \left\| D_{i,j}^l(\lambda,\varepsilon) \right\| \\
&\quad \le \frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \frac{\varepsilon^{li} \left\| D_{i,j}^l(\lambda,\varepsilon) \right\|}{\left[1 - \left|\nu_{i,j}(\lambda,\varepsilon)\right|\right]^{l+1}} \\
&\quad \le \frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \frac{2^{l+1}\varepsilon^{li} \left\| D_{i,j}^l(\lambda,\varepsilon) \right\|}{\left[-\varepsilon^i \mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))\right]^{l+1}} \\
&\le \frac{2}{N(\varepsilon)\varepsilon^i\left(-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))\right)} \sum_{l=1}^{n_{i,j}(\lambda)-1} \left[\frac{2}{-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))}\right]^l \left\| D_{i,j}(\lambda,\varepsilon) \right\|^l \\
&+ \frac{2}{N(\varepsilon)\varepsilon^i\left(-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))\right)} \sum_{l=n_{i,j}(\lambda)}^{N(\varepsilon)} \left[\frac{2}{-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))}\right]^l \left\| D_{i,j}^l(\lambda,\varepsilon) \right\| \\
&\le \frac{2}{N(\varepsilon)\varepsilon^i\left(-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))\right)} \sum_{l=1}^{n_{i,j}(\lambda)-1} \left[\frac{2}{-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))}\right]^l \left(\left\| D_{i,j}(\lambda,0) \right\| + 1\right)^l \\
&+ \frac{2}{N(\varepsilon)\varepsilon^i\left(-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))\right)}
\end{aligned}$$

$$
\begin{aligned}
&\times \sum_{l=n_{i,j}(\lambda)}^{N(\varepsilon)} \left[\frac{2}{-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))}\right]^{l} C_{i,j}(\lambda)\big(K_{i,j}(\lambda)\varepsilon\big)^{l/n^2_{i,j}(\lambda)} \\
&\le \frac{2}{N(\varepsilon)\varepsilon^{i}\big(-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))\big)} \sum_{l=1}^{n_{i,j}(\lambda)-1} \left[\frac{2\big(\|D_{i,j}(\lambda,0)\|+1\big)}{-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))}\right]^{l} \\
&+ \frac{2C_{i,j}(\lambda)}{N(\varepsilon)\varepsilon^{i}\big(-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))\big)} \sum_{l=1}^{\infty} \left[\frac{2\big(K_{i,j}(\lambda)\varepsilon\big)^{1/n^2_{i,j}(\lambda)}}{-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))}\right]^{l}.
\end{aligned}
$$

In the final expression, there are two sums, one finite and one infinite. The finite sum does not vary with ε, and is multiplied by

$$\frac{2}{N(\varepsilon)\varepsilon^{i}\big(-\mathrm{Re}\,(\bar{\lambda}\mu_{i,j}(\lambda))\big)},$$

which tends to 0 as $\varepsilon \downarrow 0$ since $N(\varepsilon)\varepsilon^{i} \to \infty$. Thus the term involving the finite sum vanishes. As for the second sum, it is a geometric series whose ratio depends on ε; in fact, this ratio is a positive constant multiplied by a fractional positive power of ε, so for ε small enough the geometric series will converge. Even stronger, the sum of the geometric series tends to 0 as $\varepsilon \downarrow 0$, and since the series is multiplied by a factor that, like the one above, also tends to 0, we see that the infinite sum term approaches 0 as $\varepsilon \downarrow 0$. Having thus bounded the original expression of interest by one that approaches 0 as $\varepsilon \downarrow 0$, the original expression, too, must go to 0. This establishes the result. □

Proposition 4.6. *If $1 \le i \le k(\lambda)$ and $N(\varepsilon)\varepsilon^{i} \to 0$ as $\varepsilon \downarrow 0$, then for $1 \le j \le n_i(\lambda)$,*

$$\lim_{\varepsilon\downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^{k}(\varepsilon)P^{*}_{i,j}(\lambda,\varepsilon) = 0.$$

Proof. Using the second estimate from Lemma 4.4, we obtain

$$
\begin{aligned}
&\left\| \frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \left[\sum_{k=l}^{N(\varepsilon)} \binom{k}{l} \nu_{i,j}^{k-l}(\lambda,\varepsilon)\right] \left[\varepsilon^{i} D_{i,j}(\lambda,\varepsilon)\right]^{l} \right\| \\
&\qquad \le \frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \left| \sum_{k=l}^{N(\varepsilon)} \binom{k}{l} \nu_{i,j}^{k-l}(\lambda,\varepsilon) \right| \varepsilon^{li} \|D^{l}_{i,j}(\lambda,\varepsilon)\| \\
&\qquad \le \frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \frac{N^{l+1}(\varepsilon)}{l!} \varepsilon^{li} \left\|D^{l}_{i,j}(\lambda,\varepsilon)\right\|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sum_{l=1}^{N(\varepsilon)} \frac{N^l(\varepsilon)\varepsilon^{li}}{l!}\left(\left\|D_{i,j}(\lambda,0)\right\|+1\right)^l \\
&\leq \sum_{l=1}^{\infty} \frac{N^l(\varepsilon)\varepsilon^{li}}{l!}\left(\left\|D_{i,j}(\lambda,0)\right\|+1\right)^l \\
&= \exp\Big(N(\varepsilon)\varepsilon^i\left(\left\|D_{i,j}(\lambda,0)\right\|+1\right)\Big)-1.
\end{aligned}
$$

Since $N(\varepsilon)\varepsilon^i \to 0$ as $\varepsilon \downarrow 0$, it follows that $N(\varepsilon)\varepsilon^i\left(\left\|D_{i,j}(\lambda,0)\right\|+1\right) \to 0$ as well, whence

$$
\exp\left(N(\varepsilon)\varepsilon^i\left(\left\|D_{i,j}(\lambda,0)\right\|+1\right)\right) \to 1.
$$

Thus the final expression in the string of inequalities goes to 0 as $\varepsilon \downarrow 0$, so the first must as well, completing the proof. $\square$

Proposition 4.7. *If* $1 \leq i \leq k(\lambda)$ *and* $N(\varepsilon)\varepsilon^i \to L$ *as* $\varepsilon \downarrow 0$, *where* $0 < L < \infty$, *then for* $1 \leq j \leq n_i(\lambda)$,

$$
\lim_{\varepsilon \downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon)P^*_{i,j}(\lambda,\varepsilon) = 0.
$$

Proof. To begin, note that since $1-\nu_{i,j}(\lambda,\varepsilon) \to 1-\lambda \neq 0$ as $\varepsilon \downarrow 0$, there is an $r > 0$ such that $\left|1-\nu_{i,j}(\lambda,\varepsilon)\right| \geq r$ for all sufficiently small positive ε. This allows us to establish the following rough estimate which we will make use of subsequently.

$$
\begin{aligned}
&\left|\frac{1-\sum_{k=0}^{l}\binom{N(\varepsilon)+1}{k}\nu_{i,j}^{N(\varepsilon)+1-k}(\lambda,\varepsilon)\left[1-\nu_{i,j}(\lambda,\varepsilon)\right]^k}{\left[1-\nu_{i,j}(\lambda,\varepsilon)\right]^{l+1}}\right| \\
&\quad\leq \frac{1+\sum_{k=0}^{l}\binom{N(\varepsilon)+1}{k}\left|\nu_{i,j}(\lambda,\varepsilon)\right|^{N(\varepsilon)+1-k}\left|1-\nu_{i,j}(\lambda,\varepsilon)\right|^k}{\left|1-\nu_{i,j}(\lambda,\varepsilon)\right|^{l+1}} \\
&\quad\leq \frac{1+\sum_{k=0}^{l}\binom{N(\varepsilon)+1}{k}2^k}{r^{l+1}} \leq \frac{1+2^l\sum_{k=0}^{l}\binom{N(\varepsilon)+1}{k}}{r^{l+1}} \\
&\quad\leq \frac{1+2^l\sum_{k=0}^{l}(N(\varepsilon)+1)^k}{r^{l+1}} \leq \frac{2^{l+1}\sum_{k=0}^{l}(N(\varepsilon)+1)^k}{r^{l+1}} \\
&\quad\leq \frac{2^{l+1}\left[(N(\varepsilon)+1)^{l+1}-1\right]}{r^{l+1}\left[(N(\varepsilon)+1)-1\right]} \leq \frac{2^{l+1}(N(\varepsilon)+1)^{l+1}}{r^{l+1}N(\varepsilon)} \\
&\quad\leq \frac{2^{2l+2}N^{l+1}(\varepsilon)}{r^{l+1}N(\varepsilon)} = \frac{2^{2l+2}N^{l}(\varepsilon)}{r^{l+1}}.
\end{aligned}
$$

Hence, using the formula from Lemma 4.4 together with the estimate just obtained and that in Lemma 4.5, we see that

$$\begin{aligned}
&\left\| \frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \left[\sum_{k=l}^{N(\varepsilon)} \binom{k}{l} \nu_{i,j}^{k-l}(\lambda,\varepsilon) \right] \left[\varepsilon^i D_{i,j}(\lambda,\varepsilon)\right]^l \right\| \\
&\quad \le \frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \left| \sum_{k=l}^{N(\varepsilon)} \binom{k}{l} \nu_{i,j}^{k-l}(\lambda,\varepsilon) \right| \varepsilon^{li} \left\| D_{i,j}^l(\lambda,\varepsilon) \right\| \\
&\quad \le \frac{1}{N(\varepsilon)} \sum_{l=1}^{N(\varepsilon)} \frac{2^{2l+2} N^l(\varepsilon)}{r^{l+1}} \varepsilon^{li} \left\| D_{i,j}^l(\lambda,\varepsilon) \right\| \\
&\quad = \frac{1}{N(\varepsilon)} \sum_{l=1}^{n_{i,j}(\lambda)-1} \frac{2^{2l+2} N^l(\varepsilon)}{r^{l+1}} \varepsilon^{li} \left(\left\| D_{i,j}^l(\lambda,0) \right\| + 1 \right)^l \\
&\quad + \frac{1}{N(\varepsilon)} \sum_{l=n_{i,j}(\lambda)}^{N(\varepsilon)} \frac{2^{2l+2} N^l(\varepsilon)}{r^{l+1}} \varepsilon^{li} C_{i,j}(\lambda) \left(K_{i,j}(\lambda)\varepsilon\right)^{l/n_{i,j}^2(\lambda)} \\
&\quad \le \frac{1}{N(\varepsilon)} \sum_{l=1}^{n_{i,j}(\lambda)-1} \frac{2^{2l+2} (N(\varepsilon)\varepsilon^i)^l}{r^{l+1}} \left(\left\| D_{i,j}^l(\lambda,0) \right\| + 1 \right)^l \\
&\quad + \frac{4C_{i,j}(\lambda)}{rN(\varepsilon)} \sum_{l=1}^{\infty} \left[\frac{4N(\varepsilon)\varepsilon^i \left(K_{i,j}(\lambda)\varepsilon\right)^{1/n_{i,j}^2(\lambda)}}{r} \right]^l .
\end{aligned}$$

In the final expression we have, as in the first proof, a finite sum and an infinite sum. Since $N(\varepsilon)\varepsilon^i \to L$ as $\varepsilon \downarrow 0$, where $0 < L < \infty$, the finite sum approaches the (finite) value

$$\sum_{l=1}^{n_{i,j}(\lambda)-1} \frac{2^{2l+2} L^l}{r^{l+1}} \left(\left\| D_{i,j}^l(\lambda,0) \right\| + 1 \right)^l .$$

As $N(\varepsilon) \uparrow \infty$, however, the term involving the finite sum goes to 0 as $\varepsilon \downarrow 0$.

The infinite sum, meanwhile, is a geometric series whose ratio goes to 0 as $\varepsilon \downarrow 0$. Thus the infinite sum also approaches 0, whence the final expression in the string of inequalities has limit 0 as $\varepsilon \downarrow 0$. It follows that the first expression has the same limit, establishing the result. □

As a consequence of these propositions, we obtain the following.

Theorem 4.5 (Krieger-Murcko). *Let λ be a unit-circle eigenvalue of P_0 not equal to* 1*. Suppose that for $1 \le i \le k(\lambda)$, $\lim_{\varepsilon\downarrow 0} N(\varepsilon)\varepsilon^i$ exists in $[0,\infty]$. Then*

$$\lim_{\varepsilon\downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*(\lambda,\varepsilon) = \lim_{\varepsilon\downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P^*_{k(\lambda)}(\lambda,\varepsilon)$$

if either of these two limits exists. If λ is completely reducible, then

$$\lim_{\varepsilon\downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon)P^*(\lambda,\varepsilon) = 0.$$

Proof. The first of these statements follows straightforwardly from Propositions 4.5, 4.6, and 4.7 once we recall the decomposition of $P^*(\lambda,\varepsilon)$ in (4.6).

Now suppose that λ is completely reducible. Then $P(\varepsilon)P^*_{k(\lambda)}(\lambda,\varepsilon) = \lambda P^*_{k(\lambda)}(\lambda,\varepsilon)$, so

$$\begin{aligned} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon)P^*_{k(\lambda)}(\lambda,\varepsilon) &= \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} \lambda^k P^*_{k(\lambda)}(\lambda,\varepsilon) \\ &= \frac{\lambda[1-\lambda^k]}{N(\varepsilon)[1-\lambda]} P^*_{k(\lambda)}(\lambda,\varepsilon) = \alpha(\varepsilon)P^*_{k(\lambda)}(\lambda,\varepsilon). \end{aligned}$$

But

$$\left|\alpha(\varepsilon)\right| = \left|\frac{\lambda[1-\lambda^k]}{N(\varepsilon)[1-\lambda]}\right| = \frac{\left|\lambda(1-\lambda^k)\right|}{N(\varepsilon)\left|1-\lambda\right|} \le \frac{2}{N(\varepsilon)\left|1-\lambda\right|} \to 0$$

as $\varepsilon \downarrow 0$, so $\alpha(\varepsilon) \to 0$ as $\varepsilon \downarrow 0$. Since also $P^*_{k(\lambda)}(\lambda,\varepsilon) \to P^*_{k(\lambda)}(\lambda,0)$, we see that

$$\lim_{\varepsilon\downarrow 0} \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon)P^*_{k(\lambda)}(\lambda,\varepsilon) = \lim_{\varepsilon\downarrow 0} \alpha(\varepsilon)P^*_{k(\lambda)}(\lambda,\varepsilon) = 0,$$

completing the proof. □

Chapter 5

Further Work

If $P(\varepsilon) = P_0 + A(\varepsilon) = P_0 + \varepsilon A_1 + \varepsilon^2 A_2 + \cdots$ is an analytically perturbed stochastic matrix and $\lambda \neq 1$ is a unit-circle eigenvalue of P_0, two main questions remain open:

(1) Is it possible for λ not to be completely reducible?

(2) If the answer to (1) is yes, what types of behavior are possible for the hybrid Cesaro limit expression in those cases?

I have little intuition as to what the answers to these questions might be, although a brief examination of the first proved suggestive. During my summer research, I made a cursory attempt to generate a perturbed stochastic matrix with -1 as a non–completely reducible eigenvalue. My approach was to take as P_0 the simplest possible 4×4 stochastic matrix with -1 as an eigenvalue of multiplicity 2 and try to find an A_1 such that the reduction process for -1 would halt after a single reduction. (Recall from Chapter 3 that the unperturbed reduced matrix in this case is equal to $P^*(-1,0)A_1P^*(-1,0)$.) With no insight as to an efficient way of trying to approach the problem, I used a brute force method that reduced the problem to solving a large system of inequalities; if a solution existed, it would yield a desired A_1. The system did not yield a solution, but the inequalities were such that if they had been relaxed slightly, a solution would have existed.

Although I was working with the simplest possible example, the way in which the system of inequalities failed to produce a solution suggested to me that there might be unknown relationships between the eigenprojection for -1 and the sign structure of the entries of A_1 that cause no example as

I was searching for to exist. With no other evidence at my disposal, I might conjecture that the answer to question (1) is no.

The difficulty in trying to prove this generally, as I see it, is the lack of any theory relating eigenvalue information about analytically perturbed matrices to the sign structure of the perturbation terms. For example, if

$$P_0 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

then the first perturbation term $A_1 = [a_{ij}]$ is very restricted in its sign structure: entries occurring where 1s are in P_0 must be nonpositive, and entries occurring where 0s are in P_0 must be nonnegative. (This follows from the requirement that $P(\varepsilon)$ remain stochastic for all sufficiently small positive ε.) It is not at all clear, though, how this sign structure can interact with the eigenprojection for -1, much less how to systematically approach this type of question.

One idea for gaining a greater understanding of the problem is to attempt to generate examples with larger matrices where there is a greater amount of freedom. Without a reasonably efficient method for going about this, however, such attempts might well fail to be illuminating and also become prohibitively complicated. This is why having at least some theoretical understanding of the situation seems fairly necessary to making any significant progress. I look forward to puzzling over this for some time to come.

Bibliography

[1] Mohammed Abbad and Jerzy A. Filar. Perturbation and stability theory for Markov control problems. *IEEE Transactions on Automatic Control*, 37(9):415–420, 1992.

KEY: AF92
ANNOTATION: Abbad and Filar study the asymptotic behavior of perturbed Markov decision problems. In particular, they investigate the existence and calculation of certain kinds of optimal strategies for these decision problems. This provides a partial motivation for studying hybrid Cesaro limits of stochastic matrices.

[2] Abraham Berman and Robert J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Academic Press, 1979.

KEY: BP79
ANNOTATION: Berman and Plemmons develop the theory of nonnegative matrices in a more sophisticated way than either [7] or [9], beginning with the idea of matrix-invariant cones. They also cover M-matrices and present applications to systems of linear equations, finite Markov chains, and input-output analysis in economics. This text, along with [9], is primarily useful for the more elementary results that can be related to stochastic matrices.

[3] François Delebecque. A reduction process for perturbed Markov chains. *SIAM Journal on Applied Mathematics*, 43(2):325–350, 1983.

KEY: fD83
ANNOTATION: Delebecque describes a reduction process for perturbed Markov chains. This reduction process gives information about the perturbed eigenvalues of the chain and

permits the determination of the perturbed chain's long-term behavior; for example, it can be applied to obtain an approximation of the perturbed chain's invariant measure. The process is used centrally in [4].

[4] Jerzy Filar, Henry A. Krieger, and Zamir Syed. Cesaro limits of analytically perturbed stochastic matrices. *Linear Algebra and its Applications*, 353:227–243, 2002.

KEY: FKS02
ANNOTATION: Filar, Krieger, and Syed characterize a hybrid Cesaro limit for analytically perturbed stochastic matrices in the case that the unperturbed matrix has no eigenvalues on the unit circle other than 1. Their characterization makes use of the reduction process in [3] via eigenprojections associated with the perturbed eigenvalues of the matrices. This article forms the basis for my work and provides key techniques that I use.

[5] G.H̃. Hardy and E.M̃. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, 5th edition, 1979.

KEY: HW79
ANNOTATION: Hardy and Wright cover a wide array of topics in number theory, from primes to continued fractions to arithmetical functions. This book is useful for its treatment of Farey series and their properties.

[6] Tosio Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, 2nd edition, 1980.

KEY: tK80
ANNOTATION: Kato devotes attention to the basic theory of linear operators on each of finite-dimensional vector spaces, Banach spaces, and Hilbert spaces. He develops perturbation theory for the finite- and infinite-dimensional cases separately, with a large focus on the latter. However it is the former, and in particular the theory concerning analytic projections, Puiseux series for perturbed eigenvalues, and the reduction process, that are important in [3], [4], and my own work.

[7] Henryk Minc. *Nonnegative Matrices*. John Wiley & Sons, 1988.

KEY: hM88
ANNOTATION: Minc develops the Perron-Frobenius theory of matrices with nonnegative entries, giving extensive coverage of the properties of eigenvalues of these matrices and also delving into more specific classes of nonnegative matrices. This book contains an inverse eigenvalue result for stochastic matrices that is essential for my work.

[8] Jason Murcko. Cesaro limits of analytically perturbed stochastic matrices with unperturbed eigenvalues on the unit circle other than 1. 2004.

KEY: jM04
ANNOTATION: This report, the result of a summer research project, formed the skeleton around which I began writing my thesis. I have retained the two examples and one proposition from the report in my thesis.

[9] Eugene Seneta. *Non-negative Matrices and Markov Chains*. Springer-Verlag, 2nd edition, 1981.

KEY: eS81
ANNOTATION: Seneta writes about both finite and infinite (countable) nonnegative matrices, though primarily the former. He briefly covers the Perron-Frobenius theory in the finite case and presents applications to, among others, Markov chains and stochastic matrices. The development of the Perron-Frobenius theory is very accessible and clear, which makes it an ideal tool for understanding some of the basic theory surrounding stochastic matrices.

[10] Zamir U. Syed. *Algorithms for Stochastic Games and Related Topics*. PhD thesis, University of Illinois at Chicago, 1999.

KEY: zS99
ANNOTATION: Syed examines stochastic games, which combine elements of matrix games and Markov decision processes. Specifically, he finds algorithms for determining optimal strategies in certain classes of stochastic games. This provides partial motivation for studying hybrid Cesaro limits of perturbed stochastic matrices, since one way of placing a valuation on strategies for these games involves Cesaro sums.