Cesaro Limits of Analytically Perturbed Stochastic Matrices

Jason Murcko

Henry Krieger, Advisor

Lesley Ward, Reader

December 10, 2004

HARVEY MUDD
COLLEGE
Department of Mathematics

Contents

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 acyclic, it is well-known that $\lim_{n\to\infty} P^n$ exists, so this limiting matrix will provide the desired long-term information about the chain's behavior.

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 towards 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 given by the projection onto the eigenspace of P for $\lambda = 1$. The Cesaro limit, sometimes referred to as the stationary matrix for P, generalizes $\lim_{n\to\infty} P^n$, as the two will evidently be 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 in actual situations, 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 we have 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 $P(\varepsilon) = P + A(\varepsilon)$, where $A(\varepsilon) \to 0$ as $\varepsilon \downarrow 0$ and $P(\varepsilon)$ remains stochastic for all sufficiently small positve 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 if this long-term behavior converges to that of the unperturbed Markov chain as the unifying parameter $\varepsilon \downarrow 0$. In other words, we may wonder if $\lim_{\varepsilon\downarrow 0} P^*(\varepsilon) = P^*$, 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

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

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. One way of doing this is

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

where $N(\varepsilon)$ takes on positive integer values and increases to ∞ as ε decreases to 0. One way of thinking about this is that for each estimate of the actual transition probabilities (i.e. each value of ε), we are only interested in the average behavior of the Markov chain up to some finite time limit (i.e. only through the first $N(\varepsilon)$ steps), but that as our estimates improve we become interested in the chain's behavior further and further out. In (4), Filar, Krieger, and Syed characterize this limit in the case that the unperturbed matrix $P = P(0)$ has no eigenvalues on the unit circle in the complex plane other than 1, in other words if P has no recurrent classes that are cyclic. In the most general case, of course, P may have cyclic recurrent classes, or equivalently eigenvalues on the unit circle other than 1. We will investigate the existence and value of the above hybrid Cesaro limit for such cases.

Chapter 2

Analytic Perturbations

In this chapter we will review some important general 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 $\mathbb{C}$), as well as results that are specific to the class of analytically perturbed stochastic matrices. In all of what follows, $M_n(\mathbb{C})$ will denote the set of complex $n \times n$ matrices.

2.1 General Results

2.1.1 Perturbed Eigenvalues and Eigenprojections

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

Definition 2.1. *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.

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 $A_1, A_2, \ldots$.

If $T(\varepsilon)$ is an analytic perturbation of T_0 (we will henceforth assume that $T(\varepsilon)$ has positive radius of convergence r_0, and will only be concerned with $T(\varepsilon)$ on the domain $D = \{ \varepsilon \mid |\varepsilon| < r_0 \}$), $T(\varepsilon)$ will have a fixed number of distinct eigenvalues except at certain "exceptional" values of ε, only a finite number of which lie in any compact set (see (5)). If λ is an eigenvalue of T_0, then $T(\varepsilon)$ will possess 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*:

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

where p_j is a positive integer. Moreover, the sum of the multiplicities of these perturbed eigenvalues in $T(\varepsilon)$ is equal to the multiplicity of λ in T_0. We refer to $\lambda_1(\varepsilon)$, $\ldots$, $\lambda_k(\varepsilon)$ as the λ-group of eigenvalues for $T(\varepsilon)$, since they all converge to λ as $\varepsilon \to 0$.

There are also useful results relating to eigenprojection matrices. In the situation above, the sum of the eigenprojections of the individual λ-group eigenvalues (note that these individual eigenprojections are only well-defined in domains which do not contain any exceptional points at which two or more of the $\lambda_j(\varepsilon)$ coincide) form an analytic perturbation of the eigenprojection for λ in T_0. This perturbed matrix, written as $P^*_\lambda(\varepsilon)$, is referred to as the *total projection* for the λ-group of eigenvalues. If the multiplicity of λ in T_0 is m, then $P^*_\lambda(\varepsilon)$ will always project onto an m-dimensional subspace of $\mathbb{C}^n$.

2.1.2 The Reduction Process

We begin here by way of an example.

Example 2.1. Consider the analytically perturbed matrix

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

T_0, the unperturbed matrix, has 2 as an eigenvalue of multiplicity 2 (in fact, $T_0 = 2I$). The perturbed eigenvalues of $T(\varepsilon)$, which both belong to the 2-group, are $\lambda_1(\varepsilon) = 2 - \varepsilon + 2\varepsilon^2$ and $\lambda_2(\varepsilon) = 2 - \varepsilon - 2\varepsilon^2$. The total projection is actually just I—there is no perturbation term for it.

Now consider the perturbed matrix

$$\widetilde{T}(\varepsilon) = \frac{1}{\varepsilon}(T(\varepsilon) - 2I)P^*_2(\varepsilon) = \frac{1}{\varepsilon}(T(\varepsilon) - 2I)I = \begin{bmatrix} -1+\varepsilon & 3\varepsilon \\ \varepsilon & -1-\varepsilon \end{bmatrix}.$$

The eigenvalues of $\widetilde{T}(\varepsilon)$ are $\tilde{\lambda}_1(\varepsilon) = -1 + 2\varepsilon$ and $\tilde{\lambda}_2(\varepsilon) = -1 - 2\varepsilon$. Note the relationship between the $\lambda_j(\varepsilon)$ and the $\tilde{\lambda}_j(\varepsilon)$: in each case,

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

In other words, the new eigenvalues $\tilde{\lambda}_j(\varepsilon)$ were obtained from the old ones by subtracting off the unperturbed part, 2, and dividing by ε.

The above is an instance of what is known as the *reduction process*. Although the above example is somewhat contrived insofar as we were able to determine the entire Puiseux series for the perturbed eigenvalues from the beginning, in the general setting when such a computation is not directly possible, we can use the reduction process to generate the coefficients in the Puiseux series one after another (subject to some limitations, as we will see).

Given an analytically perturbed matrix $T(\varepsilon)$ and a semisimple eigenvalue λ of T_0, we can reduce $T(\varepsilon)$ by computing

$$\widetilde{T}(\varepsilon) = \frac{1}{\varepsilon}(T(\varepsilon) - \lambda I)P_\lambda^*(\varepsilon). \tag{2.1}$$

If we restrict our attention to the actions of $T(\varepsilon)$ and $\widetilde{T}(\varepsilon)$ on the range of the matrix $P_\lambda^*(\varepsilon)$ (that is, the direct sum of the eigenspaces for the λ-group eigenvalues), it is not difficult to see that if $\lambda_j(\varepsilon)$ is a λ-group eigenvalue in $T(\varepsilon)$, then

$$\tilde{\lambda}_j(\varepsilon) = \frac{1}{\varepsilon}(\lambda_j(\varepsilon) - \lambda)$$

is a perturbed eigenvalue of $\widetilde{T}(\varepsilon)$ (note that this is in fact an analytically perturbed matrix as well). Conversely, if $\tilde{\lambda}_j(\varepsilon)$ is a perturbed eigenvalue of $\widetilde{T}(\varepsilon)$ (where we are still restricting our attention to the range of the matrix $P_\lambda^*(\varepsilon)$), then $\lambda_j(\varepsilon) = \lambda + \varepsilon\tilde{\lambda}_j(\varepsilon)$ is one of the perturbed λ-group eigenvalues of $T(\varepsilon)$.

It is important to note here that λ must be a semisimple eigenvalue of T_0 for equation 2.1 to work: the analytically perturbed matrix $(T(\varepsilon)-\lambda I)P_\lambda^*(\varepsilon)$ has $(T_0 - \lambda I)P_\lambda^*$ as its "constant" term, and unless λ is semisimple this will be a nonzero nilpotent matrix; in such a case, dividing $(T(\varepsilon) - \lambda I)P_\lambda^*(\varepsilon)$ by ε would *not* yield an analytically perturbed matrix, and so we would not be able to apply the nice results we have for analytically perturbed matrices.

As was alluded to, the reduction process can often be performed iteratively to yield successively more information about the Puiseux series coefficients for the perturbed eigenvalues. If λ_1 is a semisimple eigenvalue of T_0, we first reduce $T(\varepsilon)$ with respect to λ_1. Next we examine the unperturbed matrix $\widetilde{T}_0 = \widetilde{T}(0)$; if $\widetilde{T}_0$ has a semisimple eigenvalue λ_2 (again, restricting our attention to the appropriate subspace), we can reduce $\widetilde{T}_0$ with respect to λ_2; and so on. Until we reach an eigenvalue that is not semisimple for the relevant unperturbed matrix, this process can continue, yielding that the Puiseux series of some perturbed eigenvalue (or possibly eigenvalues) of $T(\varepsilon)$ begins

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

2.2 Stochastic Matrix Results

As described in chapter 1, we are primarily concerned with analytic perturbations of *stochastic* matrices, and more specifically with perturbations such that the perturbed matrix remains stochastic for all sufficiently small positive ε. We will refer to such matrices as analytically perturbed stochastic matrices, or just perturbed stochastic matrices.

The fact that the unit-circle eigenvalues of a stochastic matrix are always semisimple permits us to perform the reduction process for an analytically perturbed stochastic matrix $P(\varepsilon) = P_0 + A(\varepsilon)$ at least once for any such eigenvalue. As we will see later on, the reduction process for such eigenvalues is useful primarily insofar as it allows us to make statements about the first nonzero coefficients in the Puiseux series for the perturbed eigenvalues. To put it another way, if $P(\varepsilon) = P_0 + A(\varepsilon)$ is an analytically perturbed stochastic matrix and λ is a unit-circle eigenvalue of P_0, we are most interested in the 0-group of eigenvalues in each stage of the reduction process (following the initial reduction); once a perturbed eigenvalue splits off from this "core," we can characterize its asymptotic behavior based on the first nonzero coefficient in the Puiseux series thus produced.

One crucial result used in (4) is that, for a perturbed stochastic matrix $P(\varepsilon)$, the matrix $P(\varepsilon) - I$ can be reduced arbitrarily many times in the way described above. That is, 0 is a semisimple eigenvalue for the unperturbed matrix following each application of the reduction process. This follows from a result shown in (3), which in the given circumstances implies that the part of each reduced matrix we care about (as we reduce a matrix more and more times, we focus in on its action on subspaces of decreasing dimension) generates a Markov chain. Eventually the reduction process in this instance ceases to yield additional information: all of the perturbed eigenvalues that can split off do, and we are left with the zero matrix acting on a subspace of positive dimension (this is because 1 is still an eigenvalue of $P(\varepsilon)$, that is one of the $\lambda(\varepsilon)$ is identically equal to 1, and so for the purpose of reduction we obtain a "perturbed" eigenvalue identically equal to 0 that can never split off as the others do).

All of this implies that the first nonzero coefficient in each Puiseux series for the 1-group eigenvalues of $P(\varepsilon)$ (excluding the initial constant term of

1), if there is such a coefficient, is attached to an integer power of ε. That is, these Puiseux expansions always begin as $\lambda(\varepsilon) = 1 + c\varepsilon^k + \cdots$.

The method used in (3) cannot be reworked (at least, not straightforwardly) to give the same kind of information about arbitrary unit-circle eigenvalues. The nice results for $\lambda = 1$ arise from probabilistic interpretations of the associated eigenvectors, and these types of interpretations are lacking for the eigenvectors associated with other unit-circle eigenvalues. This is not to say, however, that the same results are *known* not to hold for other unit-circle eigenvalues. I have tried to construct examples in which the reduction process for such an eigenvalue is forced to halt due to 0's not being a semisimple eigenvalue in one of the unperturbed reduced matrices, but without luck. It would be very useful to either devise such an example or have an idea of why none might be able to exist before the conclusion of my work on this problem.

Without knowing how this issue may be resolved, I will introduce a piece of terminology to distinguish between the possible cases. Given a perturbed stochastic matrix $P(\varepsilon) = P_0 + A(\varepsilon)$ and a unit-circle eigenvalue λ of P_0, I will say that λ is *completely reducible* for $P(\varepsilon)$ if the reduction process for λ never halts because of non-semisimplicity (so 1 is completely reducible for any such $P(\varepsilon)$). More precisely, λ is completely reducible for $P(\varepsilon)$ if

1. 0 is a semisimple eigenvalue of $T_1(0)$, where

$$T_1(\varepsilon) = \frac{1}{\varepsilon}(P(\varepsilon) - \lambda I)P_\lambda^*(\varepsilon);$$

2. inductively, 0 is a semisimple eigenvalue of $T_{i+1}(0)$, where $T_{i+1}(\varepsilon)$ is obtained by performing the reduction process on $T_i(\varepsilon)$ for the eigenvalue 0.

Chapter 3

Cesaro Limit Results

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

3.1 Previous Results

Combining the information in section 2.1.1 and appendix A.1, we can 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 Cesaro limit expression as well:

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

This is essentially the approach used in (4); they showed that for any eigenvalue λ lying in the interior of the unit circle (that is, satisfying $|\lambda| < 1$),

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

regardless of how fast $N(\varepsilon) \uparrow \infty$; more importantly, though, they characterized this limit when $\lambda = 1$. In the general case, this limit *does* depend on the rate at which $N(\varepsilon) \uparrow \infty$; the expression for the limit involves certain sub-projections of $P_1^{*}(\varepsilon)$ that arise in connection with the 1-group eigenvalues of $P(\varepsilon)$ that split off during the reduction process as described in section 2.1.2.

3.2 Extending Previous Results

The previous determinations of the value of

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

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. In effect, then, to fully characterize the overall Cesaro limit is suffices to concentrate on the above limit (that is, to determine if it exists, and if so what a general expression for it will be) when λ is an eigenvalue on the unit circle other than 1. One result I obtained in this vein from my work over the summer is as follows.

Proposition 3.1. *Suppose that λ is an eigenvalue of P_0 satisfying $|\lambda| = 1$, $\lambda \neq 1$, and let $\lambda_1(\varepsilon)$, …, $\lambda_m(\varepsilon)$ be the perturbed λ-group eigenvalues, with associated eigenprojections $P^*_1(\varepsilon)$, …, $P^*_m(\varepsilon)$. If each $\lambda_i(\varepsilon)$ is a semisimple eigenvalue of $P(\varepsilon)$ and each $P^*_i(\varepsilon)$ 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,

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

By hypothesis, there exist positive constants ε_0, M_1, M_2, …, M_m such that for all $\varepsilon \in (0, \varepsilon_0)$, $P(\varepsilon)$ is stochastic and $\|P^*_i(\varepsilon)\| \leq M_i$, $i = 1, 2, \ldots, m$.

Hence, for these ε, $|\lambda_i(\varepsilon)| \leq 1$, $i = 1, 2, \ldots, m$. It follows that

$$\begin{aligned}\left\| \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P_\lambda^*(\varepsilon) \right\| &\leq \sum_{i=1}^{m} \left\| \frac{\lambda_i(\varepsilon)}{N(\varepsilon)} \frac{1 - \lambda_i^{N(\varepsilon)}(\varepsilon)}{1 - \lambda_i(\varepsilon)} P_i^*(\varepsilon) \right\| \\ &= \sum_{i=1}^{m} \frac{|\lambda_i(\varepsilon)|}{N(\varepsilon)} \frac{|1 - \lambda_i^{N(\varepsilon)}(\varepsilon)|}{|1 - \lambda_i(\varepsilon)|} \|P_i^*(\varepsilon)\| \\ &\leq \sum_{i=1}^{m} \frac{1}{N(\varepsilon)} \frac{1 + |\lambda_i(\varepsilon)|^{N(\varepsilon)}}{|1 - \lambda_i(\varepsilon)|} M_i \\ &\leq \frac{1}{N(\varepsilon)} \sum_{i=1}^{m} \frac{2M_i}{|1 - \lambda_i(\varepsilon)|},\end{aligned}$$

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 $P_i^*(\varepsilon)$ above need not be an analytic perturbation of any matrix.

From this result we can see that the only possible way the total projection for the λ-group may be able to contribute to the overall Cesaro limit is if some perturbed λ-group eigenvalue is not semisimple, or if one of the projection matrices for a perturbed λ-group eigenvalue becomes unbounded as $\varepsilon \downarrow 0$. Below I give an example of each of these two situations.

Example 3.1. For $0 < \varepsilon \leq 1$, let

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

The unperturbed matrix

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

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 is simple (algebraic multiplicity 1), while the latter is non-semisimple: its algebraic multiplicity is 2, but its geometric multiplicity is 1. The total projection $P^*_\lambda(\varepsilon)$ for the -1-group is the sum of the individual eigenprojections

$$P^*_1(\varepsilon) = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

and

$$P^*_2(\varepsilon) = \frac{1}{2} \begin{pmatrix} 1 & -1 & 0 & 0 & -1 & 1 \\ -1 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & -1 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

for $\lambda_1(\varepsilon)$ and $\lambda_2(\varepsilon)$, respectively. In addition, the nilpotent matrix associated with $\lambda_2(\varepsilon)$ is

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

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

Next we will see that

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

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

To this end, first note that

$$\begin{aligned} P^k(\varepsilon) P_\lambda^*(\varepsilon) &= P^k(\varepsilon)(P_1^*(\varepsilon) + P_2^*(\varepsilon)) \\ &= \lambda_1^k(\varepsilon) P_1^*(\varepsilon) + \lambda_2^k(\varepsilon) P_2^*(\varepsilon) + k\lambda_2^{k-1}(\varepsilon) D(\varepsilon) \\ &= (-1)^k P_1^*(\varepsilon) + (-1+\varepsilon)^k P_2^*(\varepsilon) + k\varepsilon(-1+\varepsilon)^{k-1} M. \end{aligned}$$

Hence

$$\begin{aligned} &\frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon) P_\lambda^*(\varepsilon) \\ &\qquad = \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} (-1)^k P_1^*(\varepsilon) + \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} (-1+\varepsilon)^k P_2^*(\varepsilon) \\ &\qquad\quad + \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} k\varepsilon(-1+\varepsilon)^{k-1} M. \end{aligned} \tag{3.1}$$

Let us now consider the three terms in (3.1) individually. To start,

$$\frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} (-1)^k P_1^*(\varepsilon) = \frac{-1 + (-1)^{N(\varepsilon)}}{2N(\varepsilon)} P_1^*(\varepsilon).$$

Since $P_1^*(\varepsilon)$ is a constant matrix and

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

as $\varepsilon \downarrow 0$, the first term contributes nothing to the overall limit. The second term behaves the same because $|-1+\varepsilon| \le 1$ and $P_2^*(\varepsilon)$ is also a constant matrix.

As for the third term in (3.1), we have that

$$\frac{1}{N(\varepsilon)}\sum_{k=1}^{N(\varepsilon)} k\varepsilon(-1+\varepsilon)^{k-1}M$$

$$= \frac{1-(-1+\varepsilon)^{N(\varepsilon)}[1+(2-\varepsilon)N(\varepsilon)]}{(2-\varepsilon)^2 N(\varepsilon)}\varepsilon M$$

$$= \left[\frac{\varepsilon}{(2-\varepsilon)^2 N(\varepsilon)} - \frac{\varepsilon(-1+\varepsilon)^{N(\varepsilon)}}{(2-\varepsilon)^2 N(\varepsilon)} - \frac{\varepsilon(-1+\varepsilon)^{N(\varepsilon)}}{2-\varepsilon}\right] M. \quad (3.2)$$

The numerator of each bracketed term in (3.2) approaches 0 as $\varepsilon \downarrow 0$ (in each case the magnitude of the numerator is $\leq \varepsilon$), whereas 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 (3.2) goes to 0, and since M is itself a constant matrix this implies that the third term from (3.1) tends to the zero matrix as $\varepsilon \downarrow 0$.

This shows that all three terms in (3.1) approach 0 as $\varepsilon \downarrow 0$, establishing that the overall limit is 0 as well.

Example 3.2. For $0 < \varepsilon \leq \frac{1}{2}$, let

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

The unperturbed matrix P_0 is the same as that in Example 3.1, 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 is simple. The associated eigenprojections, which sum to the total projection $P_\lambda^*(\varepsilon)$ for the -1-group, are

$$P_1^*(\varepsilon) = \frac{1}{4}\begin{pmatrix} 0 & 0 & 0 & 0 & 1-\varepsilon & -1+\varepsilon \\ 0 & 0 & 0 & 0 & -1+\varepsilon & 1-\varepsilon \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & 0 & -2 & 2 \end{pmatrix},$$

$$P_2^*(\varepsilon) = \frac{1}{4}\begin{pmatrix} 2 & -2 & -1/\varepsilon & 1/\varepsilon & -1+\varepsilon & 1-\varepsilon \\ -2 & 2 & 1/\varepsilon & -1/\varepsilon & 1-\varepsilon & -1+\varepsilon \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$P_3^*(\varepsilon) = \frac{1}{4}\begin{pmatrix} 0 & 0 & 1/\varepsilon & -1/\varepsilon & 0 & 0 \\ 0 & 0 & -1/\varepsilon & 1/\varepsilon & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & -2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Notice that $P_2^*(\varepsilon)$ and $P_3^*(\varepsilon)$ become unbounded as $\varepsilon \downarrow 0$ because of the entries that involve ε^{-1}.

As in Example 3.1, we will next see that

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

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

To begin, define new matrices

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

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

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

so that $P_2^*(\varepsilon) = M_1(\varepsilon) + M_3(\varepsilon)$ and $P_3^*(\varepsilon) = M_2 - M_3(\varepsilon)$. Then

$$\begin{aligned}
P^k(\varepsilon)P_\lambda^*(\varepsilon) &= P^k(\varepsilon)(P_1^*(\varepsilon) + P_2^*(\varepsilon) + P_3^*(\varepsilon)) \\
&= \lambda_1^k(\varepsilon)P_1^*(\varepsilon) + \lambda_2^k(\varepsilon)P_2^*(\varepsilon) + \lambda_3^k(\varepsilon)P_3^*(\varepsilon) \\
&= \lambda_1^k(\varepsilon)P_1^*(\varepsilon) + \lambda_2^k(\varepsilon)M_1(\varepsilon) + \lambda_3^k(\varepsilon)M_2 + (\lambda_2^k(\varepsilon) - \lambda_3^k(\varepsilon))M_3(\varepsilon),
\end{aligned}$$

and thus

$$\begin{aligned}
&\frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} P^k(\varepsilon)P_\lambda^*(\varepsilon) \\
&\quad = \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} \lambda_1^k(\varepsilon)P_1^*(\varepsilon) + \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} \lambda_2^k(\varepsilon)M_1(\varepsilon) \\
&\qquad + \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} \lambda_3^k(\varepsilon)M_2 + \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} (\lambda_2^k(\varepsilon) - \lambda_3^k(\varepsilon))M_3(\varepsilon).
\end{aligned} \tag{3.3}$$

The matrices $P_1^*(\varepsilon)$, $M_1(\varepsilon)$, and M_2 are all clearly bounded as $\varepsilon \downarrow 0$, and consequently the reasoning used in Proposition 3.1 can be used to show that the first three terms in (3.3) all go to 0. For the fourth term, we can make use of the fact that, for $\varepsilon \neq 0$, $\varepsilon M_3(\varepsilon)$ is a constant matrix:

$$\begin{aligned}
&\frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} (\lambda_2^k(\varepsilon) - \lambda_3^k(\varepsilon))M_3(\varepsilon) \\
&\qquad = \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} (\lambda_2(\varepsilon) - \lambda_3(\varepsilon)) \sum_{j=1}^{k} \lambda_2^{k-j}(\varepsilon)\lambda_3^{j-1}(\varepsilon)M_3(\varepsilon) \\
&\qquad = \frac{1}{N(\varepsilon)} \frac{\lambda_2(\varepsilon) - \lambda_3(\varepsilon)}{\varepsilon} \sum_{k=1}^{N(\varepsilon)} \sum_{j=1}^{k} \lambda_2^{k-j}(\varepsilon)\lambda_3^{j-1}(\varepsilon)[\varepsilon M_3(\varepsilon)] \\
&\qquad = \frac{2\varepsilon^2 - 2\varepsilon}{N(\varepsilon)} \sum_{j=1}^{N(\varepsilon)} \lambda_3^{j-1}(\varepsilon) \sum_{k=1}^{N(\varepsilon)-j+1} \lambda_2^{k-1}(\varepsilon)[\varepsilon M_3(\varepsilon)] \\
&\qquad = \frac{2\varepsilon^2 - 2\varepsilon}{N(\varepsilon)} \sum_{j=1}^{N(\varepsilon)} \lambda_3^{j-1}(\varepsilon) \frac{1 - \lambda_2^{N(\varepsilon)-j+1}(\varepsilon)}{1 - \lambda_2(\varepsilon)} [\varepsilon M_3(\varepsilon)] \\
&\qquad = \frac{1}{N(\varepsilon)} \frac{2\varepsilon^2 - 2\varepsilon}{2 - 2\varepsilon} \sum_{j=1}^{N(\varepsilon)} \lambda_3^{j-1}(\varepsilon)(1 - \lambda_2^{N(\varepsilon)-j+1}(\varepsilon))[\varepsilon M_3(\varepsilon)]
\end{aligned}$$

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

Hence

$$\begin{aligned}
&\left\| \frac{1}{N(\varepsilon)} \sum_{k=1}^{N(\varepsilon)} (\lambda_2^k(\varepsilon) - \lambda_3^k(\varepsilon)) M_3(\varepsilon) \right\| \\
&\qquad = \left\| -\frac{\varepsilon}{N(\varepsilon)} \sum_{j=1}^{N(\varepsilon)} \lambda_3^{j-1}(\varepsilon)(1 - \lambda_2^{N(\varepsilon)-j+1}(\varepsilon))[\varepsilon M_3(\varepsilon)] \right\| \\
&\qquad = \frac{\varepsilon}{N(\varepsilon)} \left| \sum_{j=1}^{N(\varepsilon)} \lambda_3^{j-1}(\varepsilon)(1 - \lambda_2^{N(\varepsilon)-j+1}(\varepsilon)) \right| \|\varepsilon M_3(\varepsilon)\| \\
&\qquad \leq \frac{\varepsilon}{N(\varepsilon)} \sum_{j=1}^{N(\varepsilon)} |\lambda_3^{j-1}(\varepsilon)(1 - \lambda_2^{N(\varepsilon)-j+1}(\varepsilon))| \|\varepsilon M_3(\varepsilon)\| \\
&\qquad \leq \frac{\varepsilon}{N(\varepsilon)} \sum_{j=1}^{N(\varepsilon)} (1 \cdot 2) \|\varepsilon M_3(\varepsilon)\| \\
&\qquad = 2\varepsilon \|\varepsilon M_3(\varepsilon)\|.
\end{aligned}$$

Since $2\varepsilon\|\varepsilon M_3(\varepsilon)\| \to 0$ as $\varepsilon \downarrow 0$, it follows that the fourth term from (3.3) goes to 0 along with the other three, implying the result.

My recent work has moved away from the type of approach exemplified by Proposition 3.1 and the examples above. That is, I have largely stopped looking at the problem in terms of whether the perturbed eigenvalues are semisimple and the individual projection matrices are bounded as $\varepsilon \downarrow 0$. Instead, pending the eigenvalue results for stochastic matrices discussed in appendix A, I am fairly confident that the methods used in (4) can be adapted to show that if $\lambda \neq 1$ is a completely reducible, unit-circle eigenvalue for $P(\varepsilon)$, then

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

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

If such an eigenvalue λ were not completely reducible, I am uncertain as to what types of behavior would be possible for the above limit expression.

This is a major reason why either finding such examples or showing that they cannot exist is important to the larger problem.

Appendix A

Matrices

In all of what follows, $M_n(\mathbb{C})$ will denote the set of all $n \times n$ matices with complex entries.

A.1 Basic Definitions

Definition A.1. *Let $T \in M_n(\mathbb{C})$, and suppose that $\lambda \in \mathbb{C}$ is an eigenvalue of T. Then we say that λ is* semisimple *(or is a semisimple eigenvalue of T) if the algebraic and geometric multiplicities of λ are equal.*

The algebraic multiplicity of an eigenvalue is always at least as large as the geometric multiplicity, so semisimplicity just means, in effect, that the eigenspace for λ, $W_\lambda = \{\, v \in \mathbb{C}^n \mid (T - \lambda I)v = 0 \,\}$, is not “deficient” (recall that the geometric multiplicity of an eigenvalue is defined to be the dimension of its associated eigenspace). Eigenspaces that are deficient, however, can be repaired by relaxing the definition as follows:

Definition A.2. *If $T \in M_n(\mathbb{C})$ and λ is an eigenvalue of T, 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 A.3. *Let $W_1, W_2, \dots, W_m$ be subspaces of $\mathbb{C}^n$. We write*

$$\mathbb{C}^n = W_1 \oplus W_2 \oplus \cdots \oplus W_m$$

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

If $\lambda_1, \dots, \lambda_m$ are the eigenvalues of a matrix $T \in M_n(\mathbb{C})$ and $W_1, \dots, W_m$ are the associated generalized eigenspaces, then it is always the case that $\mathbb{C}^n = W_1 \oplus \cdots \oplus W_m$. This statement does not hold in general, however, if we substitute in eigenspaces for generalized eigenspaces. In particular, it will fail precisely when any of the eigenvalues are 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 each individual eigenspace will not span all of $\mathbb{C}^n$.

Finally, the discussion of direct sums motivates the following:

Definition A.4. *Let $T \in M_n(\mathbb{C})$, and let the eigenvalues and generalized eigenspaces be as above. The* eigenprojection *for T associated with λ_i is the unique matrix $P_i^* \in M_n(\mathbb{C})$ such that for each $v \in \mathbb{C}^n$, $P_i^* v$ is the component of v in W_i.*

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

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

A.2 Stochastic Matrices

Definition A.5. *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.*

Let $P = [p_{ij}]$ be an $n \times n$ stochastic matrix. In what follows, $p_{ij}^{(m)}$ will denote the ijth entry of P^m.

Definition A.6. *If i and j are indices from the set $\{1, 2, \dots n\}$, we say that i has* access *to j if $p_{ij}^{(m)} > 0$ for some $m > 0$. If i has access to j and j has access to i, we say that i and j* communicate*, and write $i \leftrightarrow j$.*

It is fairly straightforward to see that, among the indices which communicate with some index, the binary relation "communicates with" is actually an equivalence relation; we may thus partition the set of all indices into communicating classes, along with an additional class containing those indices (if any) which do not communicate with any index. We can further classify the set of indices based on the following defintion.

Definition A.7. *If i is an index, we say that i is* recurrent *if, for every index j to which i has access, j has access to i. Otherwise, we say that i is* transient.

Each of the communicating classes will contain either all recurrent indices or all transient indicies; we can hence say that the classes themselves are recurrent or transient based on the classification of their indices. It is a fact that any stochastic matrix contains at least one communicating class of indices that is recurrent (see, for example, (7), p. 16). The relationships reflected in all of these classes of indices can be brought out, within a given stochastic matrix, by permuting the indices (in other words, conjugating the given matrix by a permutation matrix) so that the indices in each class are adjacent and the new matrix is lower block triangular (that is, all the positive entries occur either within or below the diagonal blocks associated with the different classes).

Eigenvalue information about P can be drawn from this index classification as well: for one thing, 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. On the other hand, with each recurrent class can be associated 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. If $k = 1$ the class is called *acyclic*, whereas if $k > 1$ the class is called *cyclic*.

Beyond this, a result of Karpelevič discussed in (6) characterizes the set of all complex numbers which are the eigenvalue of some $n \times n$ stochastic matrix, for a fixed value of n. The region is a subset of the unit circle, and intersects the unit circle at all the nth or less roots of unity; for $n \geq 4$, the boundary of the region consists of curved arcs connecting these consecutive roots of unity that are given implicitly by parametrized equations. My current work is focused on showing that these arcs approach the relevant roots of unity non-tangentially with respect to the unit circle (in other words, on showing that the angle between the unit circle and these curves at roots of unity is nonzero). I am fairly certain that this is in fact the case, but I am still in the process of working out some of the details—in particular, showing that the given parametrizations are differentiable.

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.

> 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.

> ANNOTATION: Berman and Plemmons develop the theory of nonnegative matrices in a more sophisticated way than either (6) or (7), 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 (7), are 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.

> 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 pro-

cess is used centrally in (4), and variants may prove useful in the more general problem I am considering.

[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.

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 the work I am doing and may provide insights as to possible approaches to take.

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

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.

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

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 may be useful for my work.

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

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.

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

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.