INTERMEDIATE AND LIMITING BEHAVIOR OF POWERS OF SOME CIRCULANT MATRICES

GREGORY P. TOLLISEN and TAMÁS LENGYEL

ABSTRACT. Let A be an arbitrary circulant stochastic matrix, and let $\underline{x}_0$ be a vector. An "asymptotic" canonical form is derived for A^k (as $k \to \infty$) as a tensor product of three simple matrices by employing a pseudo-invariant on sections of states for a Markov process with transition matrix A', and by analyzing how A acts on the sections, through its auxiliary polynomial. An element-wise asymptotic characterization of A^k is also given, generalizing previous results to cover both periodic and aperiodic cases. For a particular circulant stochastic matrix, identifying the intermediate stage at which fractions first appear in the sequence $\underline{x}_k = A^k \underline{x}_0$ is accomplished by utilizing congruential matrix identities and $(0, 1)$-matrices to determine the minimum 2-adic order of the coordinates of $\underline{x}_k$ through their binary expansions. Throughout, results are interpreted in the context of an arbitrary weighted average repeatedly applied simultaneously to each term of a finite sequence when read cyclically.

1. Introduction.

Consider an initial configuration of n numbers $b_0, b_1, \ldots, b_{n-1}$ distributed around a circle. Each number is then replaced by the same predetermined weighted average of (possibly itself and) its immediate and distant neighbors. More precisely, if $c_0, c_1, \ldots, c_{n-1}$ ($c_j \geq 0, \sum_{j=0}^{n-1} c_j = 1$) are the coefficients of a weighted average, we simultaneously replace each number b_l with the average $\sum_{j=0}^{n-1} c_j b_{l+j}$ where the subscripts are taken mod n. The result is one application of an averaging scheme, or "averaging around the circle" once. We wish to investigate what happens when a particular averaging scheme is applied repeatedly, beginning with a given initial configuration. The following problems [9, 11] are the prototypes of what we mean by averaging around the circle.

PROBLEM 1 Given n numbers in a circle, replace each number with the average value of itself and its clockwise neighbor for each successive pass.

$$(b_0, b_1, \ldots, b_{n-2}, b_{n-1}) \mapsto (\frac{b_0 + b_1}{2}, \frac{b_1 + b_2}{2}, \ldots, \frac{b_{n-2} + b_{n-1}}{2}, \frac{b_{n-1} + b_0}{2}).$$

2000 AMS *Mathematics Subject Classification.* 15A18, 05A15.
Key words and phrases. Circulant stochastic matrix, 0-1 matrix, congruential matrix identity.
Date: November 25, 2005

Prove that the sequence converges. Characterize the limit situation.

PROBLEM 2 Given n numbers in a circle, replace each number with the average value of itself and its clockwise second neighbor for each successive pass.

$$(b_0, b_1, \ldots, b_{n-2}, b_{n-1}) \mapsto \left(\frac{b_0 + b_2}{2}, \frac{b_1 + b_3}{2}, \ldots, \frac{b_{n-2} + b_0}{2}, \frac{b_{n-1} + b_1}{2}\right).$$

Prove that if n is odd then the sequence converges to the average of the numbers while if n is even then it eventually approaches the limit $(E, O, E, O, \ldots)$ with E and O being the averages of the numbers at even and odd positions, respectively.

PROBLEM 3 Given n numbers in a circle, replace each number with the average value of its immediate neighbors for each successive pass.

$$(b_0, b_1, \ldots, b_{n-2}, b_{n-1}) \mapsto \left(\frac{b_{n-1} + b_1}{2}, \frac{b_0 + b_2}{2}, \ldots, \frac{b_{n-3} + b_{n-1}}{2}, \frac{b_{n-2} + b_0}{2}\right).$$

Prove that if n is odd then the sequence converges to the average of the numbers. If n is even then it eventually approaches the limit cycle $\ldots \mapsto (E, O, E, O, \ldots) \mapsto (O, E, O, E, \ldots) \mapsto \ldots$ with E and O being the averages of the numbers at even and odd positions, respectively.

Many problems of this type can be found in a geometrical context such as regular generation of nested polygons [3] or at a more complex level, the Petr–Douglas–Neumann theorem on n-gons. These problems are often handled by methods presented in this paper.

In addressing problems regarding averaging around the circle, we consistently let $\underline{x}_0 = (b_0,\ b_1,\ \ldots,\ b_{n-1})$ denote the initial configuration, and let $\underline{x}_k = (b_0^{(k)}, b_1^{(k)}, \ldots, b_{n-1}^{(k)})$ be the result of k applications of the relevant averaging scheme. In Sections 2 through 4 our focus is Problem 1. In Section 2 we present a direct enumerative solution using lacunary sums, or more precisely sections of binomial coefficient sums. In matrix form, divisibility properties of these lacunary sums are applied in Section 4 to analyze the stage at which fractions first appear, when beginning with an initial configuration of integers (Theorems 2-4).

It is natural to approach these problems by describing the updating averaging step as multiplication by an appropriate $n \times n$ matrix $A = (a_{ij})$: $\underline{x}_1 = A\underline{x}_0$, and thus, $\underline{x}_k = A^k \underline{x}_0$ (where the vectors $\underline{x}_k$ are taken as column vectors). In this case, each row sum of A is equal to one. A matrix of nonnegative entries with this property is called a row-stochastic or simply a stochastic matrix. (A matrix is called column-stochastic if A' is stochastic and doubly stochastic if both A and A' are stochastic.) Further, each row of A is a simple circular shift of its first row: $a_{i,i+j} = a_{0,j}$ (where the second subscript is taken mod n). Such matrices are called circulant matrices, and therefore, $A = circ[a_{00},\ a_{01}, \ldots, a_{0,n-1}]$ is completely determined by its first row. It can be easily checked that any circulant stochastic matrix is automatically doubly stochastic and its transpose is circulant as well. In Section 3 we use circulant matrices to solve Problem 1 while in Section 5 we do the same for Problems 2 and 3.

Circulant matrices have attractive properties, many of which we will be using. Among others, it is easy to explicitly determine the eigenstructure of any circulant matrix $A = circ[c_0 = a_{0,0}, c_1 = a_{0,1}, \ldots, c_{n-1} = a_{0,n-1}]$. Let $\omega = \omega_n$ be the first nth root of unity. The eigenvalues are given by evaluating the auxiliary (sometimes also called auxiliary companion or representer) polynomial $p_A(x) = \sum_{i=0}^{n-1} c_i x^i$ at the nth roots of unity, ω^j, and thus $\lambda_j(A) = p(\omega^j)$, $j = 0, 1, \ldots, n-1$, while the eigenvectors are simply the corresponding columns of the $n \times n$ Fourier matrix $(\omega^{ij})_{i,j=0,1,\ldots,n-1}$, independent of the entries of the circulant matrix. Addition and matrix multiplication correspond to addition and multiplication $\bmod(x^n - 1)$ of their auxiliary polynomials. It follows, in particular, that circulant matrices commute. Many of these properties are easily implied by the fact that $A = p_A(E)$ with the forward shift permutation matrix $E = circ[0, 1, 0, \ldots, 0]$.

By viewing the elements of the stochastic matrix $A' = (a_{ji})$ as transition probabilities and $\underline{x}_0'$ as an initial state probability distribution, the study of the convergence of $\underline{x}_k = A^k \underline{x}_0$ as $k \to \infty$ can be modeled by Markov chains. In Section 6, we offer a brief overview of the terminology of Markov chains that we will freely use throughout the remainder of the paper. In our study the structure of the spectrum of A plays an important role, and so, we make mention of the Perron–Frobenius theorem, central to the study of matrix spectra. However, we prefer to rely on the auxiliary polynomial to provide spectral information (cf. Lemma). We note that problems that can be attacked using a Markov chain based approach have become increasingly popular in recent years, e.g., [7].

In Section 7 we introduce an approach (Theorems 6 and 7), coordinating the use of an invariant and a finely-tuned pseudo-invariant [11], that solves Problems 2, 3 and the general case (Theorem 8), and ultimately delivers an asymptotic canonical form for the powers of A and its entries (Theorem 9 and remarks). Our results generalize and extend those derived in [8]. In Section 8 we briefly touch on the problem of the first appearance of fractions for other averaging schemes.

2. First solution: direct enumeration. We set $b = \frac{1}{n}\sum_{i=0}^{n-1} b_i$ to be the arithmetic mean of the numbers $(b_0, b_1, \ldots, b_{n-1})$ and prove that the limit in Problem 1 is $\underline{b} = (b, b, \ldots, b)$. After a few preliminary exploratory computations (e.g., for the 0th position, $b_0^{(1)} = \frac{1}{2}(b_0 + b_1), b_0^{(2)} = \frac{1}{4}(b_0 + 2b_1 + b_2), b_0^{(3)} = \frac{1}{8}(b_0 + 3b_1 + 3b_2 + b_3)$, etc.), while keeping in mind the wrap-around structure, one easily conjectures that

$$b_l^{(k)} = \sum_{j=0}^{\min\{k,n-1\}} S_{j,k} b_{l+j}.$$

where

$$S_{l,k} = \sum_{t=0}^{\lfloor k/n \rfloor} \frac{1}{2^k}\binom{k}{l+nt}. \tag{2.1}$$

Induction on k will confirm the conjecture. The key to the induction step is that $b_l^{(k+1)} = \frac{1}{2}(b_l^{(k)} + b_{l+1}^{(k)})$ directly parallels the identity $\frac{1}{2^{k+1}}\binom{k+1}{l+nt} = \frac{1}{2}[\frac{1}{2^k}\binom{k}{l+nt} + \frac{1}{2^k}\binom{k}{l-1+nt}]$. A typical argument with roots of unity will show that for each l the lacunary sum $\frac{1}{2^k}\sum_t \binom{k}{l+nt}$ is asymptotically equal to $1/n$ as $k \to \infty$ [2]. This yields the limit b for $b_l^{(k)}$ as $k \to \infty$. ∎

3. Second solution: proof with circulant matrix. Here is another and perhaps, more standard approach based on stochastic matrices. Let

$$A = circ[0.5, 0.5, 0, \ldots, 0] = \begin{pmatrix} .5 & .5 & 0 & 0 & \ldots & 0 \\ 0 & .5 & .5 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ .5 & 0 & 0 & 0 & \ldots & .5 \end{pmatrix}$$

be the circulant stochastic matrix reflecting the averaging scheme in Problem 1. We want to calculate $\underline{x}_k = A^k \underline{x}_0$ as $k \to \infty$. To obtain this we need the asymptotic behavior of A^k. The auxiliary polynomial of A is $p_A(x) = 0.5 + 0.5x$, so the eigenvalues of A are $\lambda_j(A) = p_A(\omega^j) = 0.5 + 0.5\omega^j, j = 0, 1, \ldots, n-1$.

All of the eigenvalues are different since we can think of them as different points on the circle of radius 0.5 with its center at 0.5 on the x-axis. Thus, any dominant eigenvalue has magnitude $|\lambda| = 1$, implying that $\lambda = 1$ is the one and only eigenvalue of largest magnitude. Note that $\underline{v}_0 = \frac{1}{\sqrt{n}}\underline{e} = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)$ is the corresponding unit eigenvector of both A and A' (and any other circulant stochastic matrix as well). This yields the convergence of A^k to $\underline{v}_0\underline{v}_0' = \frac{1}{n}J$, thus guaranteeing $\lim_{k\to\infty} A^k \underline{x}_0 = \underline{b} = (b, b, \ldots, b)$. ∎

Interested readers may wish to consult Clark's paper [1] on the use of circulant matrices for solving other combinatorial problems. Davis' book [3] gives a comprehensive treatment of circulant matrices.

4. The first appearance of fractions. Having established that the sequence $\underline{x}_k = A^k \underline{x}_0$ converges to $\underline{b} = (b,\ b, \ldots, b)$ where b is the average of the coordinates of $\underline{x}_0$, we now shift our attention to the sequence itself. Can the sequence reach its limit in a finite number of steps, and if so, under what conditions? We address this in Theorem 1.

Now, if for example, the initial configuration $\underline{x}_0$ consists only of integers, eventually fractions must appear in the sequence. A natural question to ask is: can we say anything about when this happens? We lightly touch upon this in Theorem 1 but highlight it in Theorems 2-4. For such an initial configuration, let f denote the number of steps before the sequence "fractionizes," i.e., before a fraction is first introduced into the sequence. What can we say about the value of f?

First, note that A is nonsingular when n is odd since $p_A(\omega^j) = 0.5 + 0.5\omega^j \neq 0$ for all j. Thus, if $\underline{b}$ is reached then initially $\underline{x}_0 = \underline{b}$. On the other hand, if n is even, the null-space of A is one-dimensional since $p_A(\omega^j) = 0$ if and only if $j = \frac{n}{2}$, and is spanned by $\underline{u} = (1,\ -1,\ 1,\ -1,\ \ldots,\ 1, -1)$. Furthermore, $\underline{u}$ is not in the range of A. Thus, if $\underline{b}$ is reached then initially $\underline{x}_0 = \underline{b} + t\underline{u}$. If $\underline{x}_0$ consists only of integers then, to avoid fractions, t and b must be integers. In terms of the numbers around the circle, the limit is reached if and only if $\underline{x}_0 = (b, b, \ldots, b)$ or n is even and the initial configuration has the form $\underline{x}_0 = (E, O, \ldots, E, O)$ where $E + O = 2b$. This simple result will make a cameo appearance in the discussion preceding Theorem 4.

Since our matrices can be diagonalized, in a little more general form, we have the following

THEOREM 1. *Let A be a circulant stochastic matrix. If A is nonsingular then the limit $\underline{b}$ is reached immediately only with $\underline{x}_0 = \underline{b}$. If A is singular then the limit $\underline{b}$ is reached if and only if $\underline{x}_0 = \underline{b} + \underline{w}$, with $A\underline{w} = \underline{0}$; immediately if $\underline{w} = \underline{0}$ or in one step otherwise. Fractionizing is avoided if and only if $\underline{b}$ (and $\underline{w}$ in the singular case) consists entirely of integers.*

From now on, assume that the initial configuration $\underline{x}_0$ has only integer coordinates. In all cases not covered by the theorem, fractionizing must eventually occur. In fact, fractions are guaranteed when the range R_k of $\underline{x}_k = A^k \underline{x}_0$ (defined to be the difference between the maximum and minimum coordinates) drops below one. By observing that the minimum entry of the matrix A^n is $a = 1/2^{n-1}$ and by applying to the range R_{nk} of $\underline{x}_{nk} = A^{nk}\underline{x}_0$ the bound to be established in the proof of Theorem 5 in Section 7, $R_{nk} \leq (1 - 1/2^{n-1})^{nk} R_0$ is less than one when $k > (2^{n-1}/n) \ln R_0$. But, how tight is this bound? How far can a sequence go before fractionizing?

Clearly, if $2^K | \gcd(b_0,\ b_1,\ \ldots,\ b_{n-1})$ then fractionizing happens in more than K steps. In fact, this bound is sharp provided $K < n-1$, in the sense that f can be made to equal $K+1$ as the initial configuration $\underline{x}_0 = (2^K, 2^{K+1}, 2^{K+1}, \ldots, 2^{K+1})$ shows, by yielding a sequence which has $b_0^{(k)} = 2^{K-k}(2^{k+1}-1)$ for $0 \leq k \leq K+1$, and thus making $b_0^{(K+1)}$ a fraction while for all $k : 0 \leq k \leq K$, all values $b_l^{(k)}$ are divisible by 2^{K-k}.

On the other hand, K alone is an incomplete predictor of when fractionizing occurs. For instance, given a specific value for K, f can be made arbitrarily large if one begins with a configuration of the form $\underline{x}_0 = (2^K(2^k+1),\ 2^K(2^{k+1}+1),\ \ldots,\ 2^K(2^{k+n-1}+1))$ provided $n \geq 3$. Then fractions occur for the first time at step $K+k+1$, i.e., $f = K+k$. Nevertheless, these examples suggest the important role played by powers of 2 and further, a connection between the binary expansions of the numbers involved in the initial configuration and the number of steps required for fractionizing to occur. In fact, in the case when n is odd, the relationship is particularly simple as shown by the following theorem, and can be easily modified when n is even with an odd factor. The surprisingly complex case, which we will consider last, is when n is a nontrivial power of two.

Case 1. n is odd

THEOREM 2. *Suppose $n \geq 3$ is odd. Write each of the numbers $b_1, b_2, \ldots, b_n$ (the coordinates of $\underline{x}_0$) in base 2. Assume that l is the smallest power of two where both 0 and 1 appear among the coefficients of 2^l in the binary expansions of $b_1, b_2, \ldots, b_n$, i.e., where the binary expansions of $b_1, b_2, \ldots, b_n$ first differ. Then, it takes exactly $l+1$ steps to fractionize, i.e., $f = l$.*

Proof. Throughout this section it will be convenient to write $A = 0.5C$, where $C = circ[1, 1, 0, \ldots, 0]$. If the coordinates of $\underline{x}_0$ are not all equal, then l in the hypothesis is the maximum l for which all of the coordinates can be written in the form $b_j = c + d_j 2^l$, where c $(0 \leq c < 2^l)$ is the initial portion common to all of the coordinates. By additivity, we need only examine the behavior of $A^k(2^l\underline{d}) = 2^{l-k}C^k\underline{d}$ for the initial configuration $2^l\underline{d} = 2^l(d_0, \ldots, d_{n-1})$ and thus, fractions occur for the first time when $k = l+1$, if at least one coordinate of $C^k\underline{d}$ is odd. But, both even and odd numbers can be found among the $d_j s$. Furthermore, by induction on

k, both parities occur in $C^k \underline{d} = (d_0^{(k)}, d_1^{(k)}, \ldots, d_{n-1}^{(k)})$, for if true for $k-1$ ($k \geq 1$), then two adjacent entries $d_i^{(k-1)}$ and $d_{i+1}^{(k-1)}$ (or $d_{n-1}^{(k-1)}$ and $d_0^{(k-1)}$) have opposite parity, so $d_i^{(k)}$ is odd. However, if all of the coordinates of $C^k \underline{d}$ were odd, then the odd number of coordinates of $C^{k-1}\underline{d}$ would have to alternate parity around the circle, which is impossible. Thus, fractions are avoided only up to and including step l, i.e., $f = l$. ■

In the following, we use tensor notation to ease notational complexity and to emphasize the similarity in the block structure of many of the matrices we will encounter. Recall that the tensor (or Kronecker) product of the $m \times n$ matrix $P = (p_{ij})$ and the $m' \times n'$ matrix $Q = (q_{i'j'})$ is the $mm' \times nn'$ matrix $P \otimes Q$ formed by replacing each entry p_{ij} of matrix P with the matrix block $p_{ij}Q$, so that $(P \otimes Q)_{im'+i',\, jn'+j'} = p_{ij} q_{i'j'}$.

Case 2. n **is even with an odd factor**

An extended argument along the same lines as above can be given for the case when n is even, provided that n has at least one nontrivial odd factor. Assume $n = 2^m r$ where $r \geq 3$ is odd. Then, by inducting on m it is easy to see that $C_n^{2^m} \equiv C_r \otimes I_{2^m} \pmod 2$, where the subscripts indicate the dimensions of I and $C = circ[1, 1, 0, \ldots, 0]$. Essentially, this shows that at steps k that are multiples of 2^m, $A_n^k \underline{x}_0 = 0.5^k C_n^k \underline{x}_0$ can be partitioned into 2^m independent processes: $A_r^k \underline{b}_i = 0.5^k C_r^k \underline{b}_i$, where $\underline{b}_i = (b_i, b_{i+2^m}, b_{i+2\cdot 2^m} \ldots, b_{i+(r-1)2^m})$ is the ith section of the original configuration $\underline{x}_0$, each being an independent odd case covered by Theorem 2. After a little effort reintegrating the results for each of the sections, one arrives at the following

THEOREM 3. *Suppose $n = 2^m r$ is even with $r \geq 3$ odd. For each $i = 0, 1, \ldots, 2^m - 1$, we write each of the numbers $b_i, b_{i+2^m}, b_{i+2\cdot 2^m} \ldots, b_{i+(r-1)2^m}$ (the ith section of $\underline{x}_0$) in base 2. Let $l_i \geq 0$ be the smallest power of two for which both 0 and 1 can be found as coefficients of 2^{l_i} in the binary expansions of these numbers. In other words, l_i is the power of two corresponding to the first position where the binary expansions of $b_i, b_{i+2^m}, b_{i+2\cdot 2^m} \ldots, b_{i+(r-1)2^m}$ do not all agree. Then, $f = \min_i l_i$.*

Case 3. n **is a nontrivial power of 2**

The situation when n is a nontrivial power of 2 is unusual because in this case and only in this case, powers of $C = circ[1,\ 1,\ 0, \ldots, 0]$ can be found with all even entries, which can extend the number of steps before fractionizing occurs. In fact, a power of C can be found with all entries divisible by any given power of 2. To take this into account, we will need the following

Facts about powers of C: If $C = circ[1,1,0,\ldots,0]$ and $E = circ[0,1,0,\ldots,0] = C - I$ are $n \times n$ where $n = 2^m$, then powers of C that are multiples of 2^{m-1} take on a particularly simple pattern.

F1: 2 does not divide $C^t, 0 \leq t < 2^{m-1}$, and 2^{s-1} is the highest power of 2 dividing $C^{s \cdot 2^{m-1}+t}$, $s \geq 1,\ 0 \leq t < 2^{m-1}$

F2:
$$C^{2^{m-1}} \equiv \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \otimes I_{2^{m-2}} \equiv \begin{pmatrix} I & I \\ I & I \end{pmatrix} \bmod 2$$

F3:
$$\begin{aligned} 2^{-2r+1} C^{(2r)2^{m-1}} &\equiv \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \otimes I_{2^{m-2}} \equiv \\ &\equiv \begin{pmatrix} I & I \\ I & I \end{pmatrix} \bmod 2,\ r \geq 1 \end{aligned}$$

F4:
$$\begin{aligned} 2^{-2r} C^{(2r+1)2^{m-1}} &\equiv \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \otimes I_{2^{m-2}} \equiv \\ &\equiv E^{2^{m-2}} \begin{pmatrix} I & I \\ I & I \end{pmatrix} \bmod 2,\ r \geq 1 \end{aligned}$$

Note that the right hand sides of F3 and F4 do not depend on r.

These matrix facts are translations of results on the divisibility properties of lacunary sums of binomial coefficients derived in [10] appearing as coefficients in the auxiliary polynomial $p_{C^k}(x) \equiv (1+x)^k \mod (x^n - 1)$ for C^k. (Fact F1 is included as background for F2-F4. Also repeatedly used will be the simple observation that with integer matrices A and B : $\gcd(A)\gcd(B) | \gcd(AB)$ with $\gcd(D)$ denoting the greatest common divisor of all entries in matrix D.)

Suppose each coordinate b_i of the initial configuration $\underline{x}_0 = (b_0, b_1, \ldots, b_{n-1})$ is written in its binary form $b_i = \beta_{i,0} + \beta_{i,1}2 + \beta_{i,2}2^2 +$

$\cdots + \beta_{i,d}2^d$. Then $\underline{x}_0 = \underline{\beta}_0 + \underline{\beta}_1 2 + \underline{\beta}_2 2^2 + \cdots + \underline{\beta}_d 2^d$, where each $\underline{\beta}_j = (\beta_{0,j}, \beta_{1,j}, \ldots, \beta_{n-1,j})$ is a $(0,1)$-vector. We consider each $\underline{\beta}_j 2^j \neq \underline{0}$ separately as an *elementary* initial configuration and determine when it fractionizes. If any one of the initial configurations $\underline{\beta}_j 2^j$ fractionizes at step k before the rest, then by additivity, the sum $\underline{x}_0$ of the elementary initial configurations will fractionize at that same step k.

An analogy might be helpful here and throughout. We can imagine a *race* with each initial configuration $\underline{\beta}_j 2^j$ as a contestant initially located in binary position j corresponding to the power 2^j, the goal of the race being the first to fractionize. At each time step, the averaging scheme is applied. After k applications the contestant has reached the binary position s where 2^s is the greatest power of 2 that divides all components of $A^k\left(\underline{\beta}_j 2^j\right)$. If one contestant passes the "finishing line" before any of the others (i.e., passes beyond the 0th or unit binary position) then the sum $\underline{x}_0$ fractionizes at that same precise time step. (Ties will be considered later.)

Now, define the rotational period (or simply the period) of $\underline{\beta}_j$ to be the least $p > 0$ such that $\beta_{i+p,j} = \beta_{i,j}$ for all i where the first subscript is taken mod n. Then $p|n = 2^m$ and p is a power of 2. If $p = n$ we say that $\underline{\beta}_j$ is rotationally asymmetric. The period p determines how slowly the process of repeated averagings fractionizes the initial configuration $\underline{\beta}_j 2^j$. If $p = 1$ then $\underline{\beta}_j 2^j$ is constant and at no step k will $A^k\left(\underline{\beta}_j 2^j\right)$ contain a fraction, so $\underline{\beta}_j 2^j$ will not fractionize. If $p = 2$ then we are in the context of Theorem 1 and $\underline{\beta}_j 2^j$ fractionizes in one step if $j = 0$ (i.e., $f = 0$). Otherwise, $\underline{\beta}_j 2^j$ never fractionizes.

If $2 < p < n$, then the initial configuration $\underline{\beta}_j 2^j$ of size n can be reduced to an initial configuration of size p because the coordinates of $\underline{\beta}_j = (1,\ 1,\ \ldots,\ 1)_{n/p} \otimes (\beta_{0,j}, \beta_{1,j}, \ldots, \beta_{p-1,j})$ are simply the first p coordinates of $\underline{\beta}_j$ repeated n/p times. The result of k successive averages of $A_n^k\left(\underline{\beta}_j 2^j\right) = (1, 1, \ldots, 1)_{n/p} \otimes A_p^k\left((\beta_{0,j}, \beta_{1,j}, \ldots, \beta_{p-1,j})2^j\right)$, is the same as k successive averages of the initial configuration $(\beta_{0,j}, \beta_{1,j}, \ldots, \beta_{p-1,j})$ around a circle of size p. Thus, without loss of generality, we can reduce $\underline{\beta}_j$ to its first p coordinates and reduce the dimension n to match the period p to get the following final case.

When $n = 2^m = p$, note that $C^{2^{m-1}}\underline{\beta}_j \not\equiv \underline{0} \pmod 2$, for

$\begin{pmatrix} I & I \\ I & I \end{pmatrix} \underline{\beta}_j \equiv \underline{0} \pmod 2$ (cf. F2) would imply $p | 2^{m-1}$, contrary to the case we are considering. On the other hand, $C^{2^m} \underline{\beta}_j \equiv \underline{0} \pmod 2$ since $2 | C^{2 \cdot 2^{m-1}}$ (cf. F3 with $r = 1$). So, let $0 < \varphi \leq 2^{m-1}$ be least such that $C^{2^{m-1}+\varphi} \underline{\beta}_j \equiv \underline{0} \pmod 2$. Note that this implies $\begin{pmatrix} I & I \\ I & I \end{pmatrix} C^{\varphi} \underline{\beta}_j \equiv \underline{0} \pmod 2$ but $\begin{pmatrix} I & I \\ I & I \end{pmatrix} C^{\varphi-1} \underline{\beta}_j \not\equiv \underline{0} \pmod 2$ (cf. F2). Now for $k \geq \varphi$, we claim that the highest power of 2 that divides (all entries of) $C^k \underline{\beta}_j$ is $2^{\left\lfloor \frac{k-\varphi}{2^{m-1}} \right\rfloor}$. To show this, it suffices to show that for the extremes of the range $s \cdot 2^{m-1} + \varphi \leq k < (s+1) 2^{m-1} + \varphi$, ($s \geq 1$, $s = 0$ being obvious) the highest power of 2 dividing $C^k \underline{\beta}_j$ is 2^s.

But for the upper end $C^{(s+1)2^{m-1}+\varphi-1} \underline{\beta}_j = C^{(s+1)2^{m-1}} C^{\varphi-1} \underline{\beta}_j$ and $2^{-s} C^{(s+1)2^{m-1}}$ is a matrix of integers congruent to $\begin{pmatrix} I & I \\ I & I \end{pmatrix}$ or $E^{2^{m-2}} \begin{pmatrix} I & I \\ I & I \end{pmatrix} \pmod 2$ (cf. F3 and F4). Thus $2^{-s} C^{(s+1)2^{m-1}} C^{\varphi-1} \underline{\beta}_j \not\equiv \underline{0} \pmod 2$, so $C^{(s+1)2^{m-1}+\varphi-1} \underline{\beta}_j$ is divisible by 2^s and no higher power of 2.

The preceding directly implies that at the lower end $C^{s \cdot 2^{m-1}+\varphi} \underline{\beta}_j = C^{s \cdot 2^{m-1}} C^{\varphi} \underline{\beta}_j$ is not divisible by any power of 2 beyond 2^s. On the other hand, $2^{-s+1} C^{s \cdot 2^{m-1}}$ is a matrix of integers congruent to $\begin{pmatrix} I & I \\ I & I \end{pmatrix}$ or $E^{2^{m-2}} \begin{pmatrix} I & I \\ I & I \end{pmatrix} \pmod 2$ (cf. F3 and F4). Thus $2^{-s+1} C^{s \cdot 2^{m-1}} C^{\varphi} \underline{\beta}_j \equiv \underline{0} \pmod 2$, so $C^{s \cdot 2^{m-1}+\varphi} \underline{\beta}_j$ is divisible by 2^s. We are now ready to establish

THEOREM 4. *Suppose $\underline{\beta}_j$, $j \geq 0$, is a rotationally asymmetric $(0,1)$-vector of size (and period) p. Then $A^k(\underline{\beta}_j 2^j)$, i.e., the averaging scheme A applied k times to the elementary initial configuration $\underline{\beta}_j 2^j$, avoids fractions only up to and including step $k = f$ where*

$$f = \begin{cases} j, & \text{if } j < \varphi, \\ \left\lfloor \frac{pj - 2\varphi}{p-2} \right\rfloor, & \text{if } j \geq \varphi. \end{cases}$$

Proof. This is clear for $j < \varphi$, for the initial configuration $\underline{\beta}_j 2^j$ avoids fractionizing in $k < 2^{m-1} + \varphi$ averages as long as $A^k\left(\underline{\beta}_j 2^j\right) = 2^{j-k} C^k \underline{\beta}_j$ avoids fractions, which occurs for the last time when $k = j$. On the other hand, if $j \geq \varphi$, then one can average $k \geq \varphi$ times before the initial configuration $\underline{\beta}_j 2^j$ fractionizes. We calculate

$$A^k\left(\underline{\beta}_j 2^j\right) = 2^{j-k+\left\lfloor \frac{k-\varphi}{2^{m-1}} \right\rfloor} \left[2^{-\left\lfloor \frac{k-\varphi}{2^{m-1}} \right\rfloor} C^k \underline{\beta}_j\right] \tag{4.1}$$

where the expression in the square brackets has integer entries of both parities. The maximum k for which $j - k + \left\lfloor \frac{k-\varphi}{2^{m-1}} \right\rfloor \geq 0$ can be checked to be $k = \left\lfloor \frac{2^{m-1} j - \varphi}{2^{m-1} - 1} \right\rfloor = \left\lfloor \frac{pj - 2\varphi}{p-2} \right\rfloor$. ∎

One way to interpret this result is to rewrite the formula for

$$f = \begin{cases} \frac{j}{1}, & \text{if } j < \varphi, \\ \left\lfloor \frac{\varphi}{1} + \frac{j-\varphi}{1-\frac{2}{p}} \right\rfloor, & \text{if } j \geq \varphi, \end{cases}$$

and consider the numerators as distances and denominators as "average" velocities in the racing context. For a contestant's configuration $\underline{\beta}_j 2^j$ initially located in the jth binary position where $j < \varphi$, $k \leq \varphi$ applications of the averaging scheme give the result $A^k\left(\underline{\beta}_j 2^j\right) = 2^{j-k} C^k \underline{\beta}_j$. Thus, for each time step, the averaging scheme is applied once, k increases by 1, the common power of 2 decreases by 1, and the contestant is one binary position closer to the goal. Thus, the contestant approaches the goal with velocity 1 from the initial binary position j. On the other hand, if the contestant begins in binary position $j \geq \varphi$, for the first φ applications of the averaging scheme, the contestant's position decreases with velocity 1 as before. But for $k \geq \varphi$ applications of the averaging scheme we have the more complicated result (4.1). Again, for each time step, k increases by 1, decreasing the common power of 2 by 1 *except* when $k \equiv \varphi - 1 \pmod{2^{m-1}}$ where no decrease occurs. Thus, in 2^{m-1} time steps, the contestant advances $2^{m-1} - 1$ binary positions with velocity $\frac{2^{m-1}-1}{2^{m-1}} = 1 - \frac{2}{p}$. Thus, to reach the finish line, essentially, the contestant advances with velocity 1 through the first φ time steps and with velocity $1 - \frac{2}{p}$ through the final $j - \varphi$ from initial binary position j to final binary position 0 immediately before fractionizing.

We conclude that from their various initial positions, no contestant approaches fractionizing with a velocity greater than 1; that except for contestants with periods $p = 1$ and 2, who initially or after one step, respectively "lose strength" and stop, all other contestants begin with the same velocity 1 through their initial stretch of length φ, but beyond the initial stretch, immediately decrease their velocities below 1, to velocities completely determined by their period, with smaller velocities corresponding to smaller periods.

We include here a simple algorithm for determining φ for a given $(0,1)$-vector $\underline{z}$ of size and period $n = p = 2^m$. The strategy is to search for powers of 2, beginning with 2^{m-1} and working downwards that, when collectively used as exponents for C, the resulting product will not annihilate $\underline{z}$ when taken mod 2. In other words, begin with $\epsilon_{m-1} = 1$ and select $\epsilon_{m-i} = 0$ or 1, $i = 2, 3, 4, \ldots, m$, with as many $\epsilon_{m-i}s$ as possible equaling 1 while still maintaining $C^{2^{m-1}} C^{\epsilon_{m-2} 2^{m-2}} C^{\epsilon_{m-2} 2^{m-2}} \ldots C^{\epsilon_0} \underline{z} \not\equiv \underline{0}$ (mod 2). The resulting power of C will be $2^{m-1} + \varphi - 1$, the maximum power of C that does not annihilate $\underline{z}$. Then, $\varphi = \epsilon_{m-2} 2^{m-2} + \epsilon_{m-3} 2^{m-3} + \cdots + \epsilon_0 + 1$.

The following algorithm systematically determines the $\epsilon_{m-i}s$ while avoiding repetitions due to progressively smaller periods.

ALGORITHM. The strategy is:

1. begin with $\underline{z}_0 = \underline{z}$ of size 2^m and let $\epsilon_{m-1} = 1$;
2. having found $\underline{z}_i$ of size 2^{m-i}: if the period of $\underline{z}_i$ is 2^{m-i} then let $\epsilon_{m-i-1} = 1$ and let $\underline{z}_{i+1} \equiv (z_0 + z_{2^{m-i-1}},\ z_1 + z_{2^{m-i-1}+1},\ z_2 + z_{2^{m-i-1}+2},\ \ldots,\ z_{2^{m-i-1}-1} + z_{2^{m-i}-1})$ (mod 2) of size 2^{m-i-1}; otherwise, let $\epsilon_{m-i-1} = 0$ and let $\underline{z}_{i+1} = (z_0,\ z_1,\ \ldots,\ z_{2^{m-i-1}})$ of size 2^{m-i-1};
3. end when ϵ_0 is determined.

Then the least k for which $C^k \underline{z} \equiv \underline{0}$ (mod 2) is $k = \epsilon_{m-1} 2^{m-1} + \epsilon_{m-2} 2^{m-2} + \cdots + \epsilon_0 + 1$ and $\varphi = \epsilon_{m-2} 2^{m-2} + \cdots + \epsilon_0 + 1$.

For example, suppose $\underline{z}$ = (1,1,0,0,0,1,1,0). Then, $n = 8$ and $m = 3$, and $\underline{z}_0 = (1, 1, 0, 0, 0, 1, 1, 0)$ with $\epsilon_2 = 1$, $\underline{z}_1 = (1, 0, 1, 0)$ with $\epsilon_1 = 0$, and $\underline{z}_2 = (1, 0)$ with $\epsilon_0 = 1$, so $k = \left[1(2^2) + 0(2) + 1\right] + 1 = 6$ and $\varphi = [0(2) + 1] + 1 = 2$.

We conclude with a few observations and an example. When analyzing the number of steps before the initial configuration $\underline{x}_0 = \underline{\beta}_0 + \underline{\beta}_1 2 + \cdots + \underline{\beta}_d 2^d$ fractionizes, we divide the individual elementary initial configurations $\underline{\beta}_j 2^j$ into groups according to period. Those with period $p = 0$ can be immediately disregarded and those with period $p = 1$ can as well, unless $j = 0$, with fractionizing occurring in one step. For each of the other periods present, only the elementary initial configuration with the smallest j in the group corresponding to the period need be considered as a candidate for winning. The value of f for each candidate can be calculated, and if one value is strictly less than all others, then this minimum is the value of f for the initial configuration $\underline{x}_0$. Otherwise, there is a tie, when two or more candidates fractionize at the same step. Then it is possible that $A^{f+1}\left(\underline{\beta}_j 2^j\right)$ contain fractions for two or more individual candidates $\underline{\beta}_j 2^j$, while the sum $A^{f+1}\underline{x}_0 = A^{f+1}\underline{\beta}_0 + A^{f+1}\left(\underline{\beta}_1 2\right) + \cdots + A^{f+1}\left(\underline{\beta}_d 2^d\right)$ contains only integers, thus delaying the fractionizing step. However, because of difference in velocity, this cannot continue beyond the point where one summand leads the rest in the sense that it contains a fraction whose denominator is a power of 2 strictly greater than the denominators of the other summands. Then, the exponent for this dominant power of 2 can be used to trace back to an upper bound on the step before fractionizing, since each averaging step can increase the denominator by at most one power of 2.

EXAMPLE. Suppose we begin with the numbers:

$$\begin{aligned}
b_0 &= 481 = 00111100001_2\\
b_1 &= 1473 = 10111000001_2\\
b_2 &= 161 = 00010100001_2\\
b_3 &= 1985 = 11111000001_2\\
b_4 &= 417 = 00110100001_2\\
b_5 &= 1153 = 10010000001_2\\
b_6 &= 481 = 00111100001_2\\
b_7 &= 1665 = 11010000001_2
\end{aligned}$$

We wish to find the last step before fractionizing. We break up the columns according to their individual periodicities:

period 1: columns corresponding to the $2^0,\ 2^1,\ 2^2,\ 2^3,\ 2^4,\ 2^7$ places.
period 2: columns corresponding to the $2^5,\ 2^{10}$ places.
period 4: column corresponding to the 2^9 place.
period 8: columns corresponding to the $2^6,\ 2^8$ places.

Only that column in each period category corresponding to the lowest power of 2 need be considered. Those of period 1 are immediately thrown out. Their velocity is 0. Those of period 2 will not fractionize. The smallest power of 2 in this category is not 2^0, so these will not fractionize. We need only consider the columns corresponding to 2^9 and 2^6 (as 2^6 is the frontrunner for the velocity corresponding to $p = 8$). For 2^9, $p = 4 = 2^2,\ m = 2,\ j = 9$ and $\underline{\beta}_9 = (0,0,0,1)$. Now, $\underline{z}_0 = (0,0,0,1)$ has period 4 so $\epsilon_1 = 1$ and $\underline{z}_1 \equiv (0+0, 0+1) \equiv (0,1) \pmod 2$. Further, $\underline{z}_1 = (0,1)$ has period 2, so $\epsilon_0 = 1$. Thus $\varphi = \epsilon_0 2^0 + 1 = 2$ and

$f = \left\lfloor \frac{pj-2\varphi}{p-2} \right\rfloor = \left\lfloor \frac{(4)(9)-2(2)}{4-2} \right\rfloor = 16$. For 2^6, $p = 8 = 2^3$, $m = 3$, $j = 6$ and $\underline{\beta}_6 = (1,1,0,1,0,0,1,0)$. Now, $\underline{z}_0 = (1,1,0,1,0,0,1,0)$ has period 8, so $\epsilon_2 = 1$ and $\underline{z}_1 \equiv (1+0, 1+0, 0+1, 1+0) \equiv (1,1,1,1) \pmod 2$. Next, $\underline{z}_1$ doesn't have period 4, so $\epsilon_1 = 0$ and $\underline{z}_2 = (1,1)$. Finally, $\underline{z}_2$ doesn't have period 2, so $\epsilon_0 = 0$. Thus $\varphi = \epsilon_1 2 + \epsilon_0 + 1 = 1$ and $f = \left\lfloor \frac{pj-2\varphi}{p-2} \right\rfloor = \left\lfloor \frac{(8)(6)-2(1)}{8-2} \right\rfloor = 7$. Consequently, for the original numbers, $f = 7$, the minimum between the two possibilities.

In the last section, we will return to these issues, where we briefly comment on other averaging schemes.

5. Solutions to Problems 2 and 3 by circulant matrices. We now solve Problems 2 and 3 with our eyes open to structural characteristics that may show up in general. First we note that, with a bit of ingenuity, both problems can be reduced to Problem 1. If n is odd, the reduction is exact for Problem 2 after reordering the positions so that the even positions come before the odd positions (i.e., positions $0, 2, 4, \ldots, 1, 3, 5, \ldots$). Then Problem 3 is reduced to Problem 2 by following each application of the averaging scheme with a clockwise shift of the original labeling. If n is even, the positions are split into two groups: the even positions and the odd positions. In Problem 2, it is clear that the averaging scheme acts on each group independently, while in Problem 3, one application of the averaging scheme acts on each group to exclusively determine the values in the other group, resulting in alternating behavior between the groups. Either way, the problems are reduced to two independent instances of Problem 1. (Therefore, the fractionizing questions for all instances of Problems 2 and 3 can be answered directly by the results in Section 4.)

At this point it is difficult to see whether or not the splitting strategy used above applies to any given averaging process around the circle, and if it does, how it is to be accomplished. Perhaps the eigenvalue route used in Section 3 will be more helpful.

The circulant matrix corresponding to the averaging scheme used in Problem 2 is $A = circ[0.5, 0, 0.5, 0, \ldots, 0]$ where the auxiliary polynomial $p_A(x) = 0.5 + 0.5x^2$ determines the eigenvalues to be $\lambda_j(A) = 0.5 + 0.5\omega^{2j}$ with $|\lambda_j(A)| \leq 1$. If n is odd then $|\lambda_j(A)| = 1$ precisely when $j = 0$, giving the eigenvalue $\lambda_0(A) = 1$ of multiplicity one, and the problem can be finished as in Section 3.

When n is even, $|\lambda_j(A)| = 1$ precisely when $\omega^{2j} = 1$, i.e., when $j = 0$ and $\frac{n}{2}$. Both give the single dominant eigenvalue $\lambda = 1$ with multiplicity 2. One easily checks that its two-dimensional eigenspace is spanned by the orthonormal eigenvectors $\underline{v}_0 = \frac{1}{\sqrt{n}}\underline{e} = \frac{1}{\sqrt{n}}(1,1,\ldots,1,1)$ and $\underline{v}_1 = \frac{1}{\sqrt{n}}\underline{u} = \frac{1}{\sqrt{n}}(1,-1,\ldots,1,-1)$. Therefore, we get $A^k\underline{x}_0 \approx \left(1^k\underline{v}_0\underline{v}_0' + 1^k\underline{v}_1\underline{v}_1'\right)\underline{x}_0 = \frac{1}{n}circ[2,0,\ldots,2,0]\,\underline{x}_0 = (E,O,\ldots,E,O)$ (in the sense that the difference approaches $\underline{0}$ as $k \to \infty$), where $E = \frac{2}{n}(b_0 + b_2 + \cdots + b_{n-2})$ and $O = \frac{2}{n}(b_1 + b_3 + \ldots + b_{n-1})$.

For Problem 3 the matrix $A = A'$ has auxiliary polynomial $p_A(x) = 0.5x + 0.5x^{n-1} = x(0.5 + 0.5x^{n-2})$ and eigenvalues $\lambda_j(A) = \omega^j(0.5 + 0.5\omega^{(n-2)j})$. We note again that $|\lambda_j(A)| \le 1$. The case with n odd is identical to that above. When n is even, $|\lambda_j(A)| = 1$ precisely when the second factor $0.5 + 0.5\omega^{(n-2)j} = 0.5 + 0.5\omega^{-2j}$ has magnitude one, i.e., when $j = 0$ and $\frac{n}{2}$. But here, the first factor gives us two different eigenvalues: $\lambda_0(A) = 1$ and $\lambda_{n/2}(A) = -1$, each with eigenspace spanned by $\underline{v}_0$ and $\underline{v}_1$, respectively. Then $A^k\underline{x}_0 \approx \left(1^k\underline{v}_0\underline{v}_0' + (-1)^k\underline{v}_1\underline{v}_1'\right)\underline{x}_0 = \frac{1}{n}(c,c,c,c,\ldots,c,c) + (-1)^k\frac{1}{n}(d,-d,d,-d,\ldots,d,-d)$ with $c = \sum_{i=0}^{n-1} b_i$ and $d = \sum_{i=0}^{n-1}(-1)^i b_i$. This yields $\lim_{k\to\infty} A^{2k}\underline{x}_0 = (E,O,E,O,\ldots,E,O)$ and $\lim_{k\to\infty} A^{2k+1}\underline{x}_0 = (O,E,O,E,\ldots,O,E)$, with E and O defined as above.

From the solutions to Problems 1, 2 and 3, we see differing, though related phenomena. Positions around the circle partition into natural subsets (Problems 2 and 3), or are acted on as a whole (Problem 1) by the averaging scheme. The result on each subset (or the whole) is to reach a single common value, or to approach a cycle of values, and the values are always averages of sections (or the whole) of the initial configuration. At this point, the reader familiar with finite Markov processes will readily recognize these patterns as limiting and asymptotically periodic behavior on recurrent classes. Thus, in the next section, we review the ideas and vocabulary of Markov chains.

In passing, we note that other similar problems can be solved by turning to eigenvectors as above. Here we mention only a particular one [11, Problem 3.2.8]: There are n points on a circle, and each point is given a number which is equal to the average of the numbers of its two nearest neighbors. Show that all of the numbers must be equal. This problem can be easily solved by observing that $A\underline{x}_0 = \underline{x}_0$, with the setting of Problem 3, the eigenvalue 1 is simple and $\underline{e}$ is a corresponding eigenvector. Another approach would be to apply the extreme principle [11] to the coordinates of $\underline{x}_0$.

6. Markov chains. We now wish to study the limiting behavior of averaging schemes around the circle in general. It should be clear at this point that, no matter what averaging scheme we consider, the fundamental task is to characterize how powers of the corresponding circulant stochastic matrix behave asymptotically, and that the dominant eigenvalues in the spectrum of the matrix play a central role in achieving this task. Fortunately, this is ground already well covered in the context of finite Markov chains, and we will liberally use ideas and terms from that field, which in this section we intend to summarize for the benefit of the reader. We first consider how the states of a Markov chain are partitioned, classified and characterized by its transition matrix. Then we consider the eigenvalues of a transition matrix. Throughout, we draw conclusions within the specific context of circulant stochastic matrices.

A technicality: In studying the limiting behavior of $\underline{x}_k = A^k \underline{x}_0$ we use $M = A'$ as a transition matrix to fit the form $\underline{x}'_k = \underline{x}'_0 M^k$ commonly used in Markov chains. However, that A (and A') is circulant ultimately makes this difference irrelevant to us, as will be explained below.

Suppose $M = (m_{ij})$ is an $n \times n$ stochastic matrix. The matrix M can be interpreted as a transition matrix for a finite Markov chain with n states, where m_{ij} is the probability of passing from state i to state j in one step. Associated with the transition matrix M is its dependency graph G, a digraph whose vertices $V = \{0, 1, 2, \ldots, n-1\}$ are the states, and which includes each edge $i \to j$ if and only if $m_{ij} > 0$. We say that states i and j communicate with each other if there is a directed path in G from state i to state j, and one from state j to state i. With the notation $M^k = (m_{ij}^{(k)})$ this means that there exist k_1 and k_2 so that $m_{ij}^{(k_1)} > 0$ and $m_{ji}^{(k_2)} > 0$. This defines an equivalence relation that partitions the vertices into classes. Graphically, each class is a strongly connected subgraph of G. If a Markov chain (and its transition matrix) has only one class, it is called irreducible.

A state i is recurrent if the probability of returning to that state in a finite number of steps is positive, i.e., if $m_{ii}^{(k)} > 0$ for some k. Otherwise, i is transient. Both are class properties, so each class can be termed either recurrent or transient. There must be at least one recurrent class, and every recurrent class is closed, in the sense that every edge originating in the recurrent class must terminate there as well. Our context is less complicated. The circulant matrix A' treats the states (i.e., positions around the circle) in a rotationally symmetric manner. This guarantees that if A' is not irreducible (i.e., has more than one class) then all of the classes are of the same size, are recurrent exclusively and partition the graph into strongly connected sets that are mutually disconnected. In addition, the states within each class are regularly distributed around the circle (i.e., each class is rotationally symmetric), and any two classes differ only by a rotation. Furthermore, the symmetry of the communication relation implies that the classes of A and A' are the same. All of these facts reveal further structure within A. Indeed, through conjugation with an appropriate permutation matrix, A is equivalent to a block diagonal matrix with as many equal blocks as there are classes.

State i is periodic with parameter p if all closed paths through i have a length that is a multiple of p, i.e., if $m_{ii}^{(k)} > 0$ implies $p|k$. The maximum such p is called the period of state i. The period is also a class property, so a class and A as well (if A is irreducible) are said to have period p. If $p = 1$ then the state and its class are called aperiodic. A recurrent and aperiodic state (or class) is also called an ergodic state (or class). A class with period $p > 1$ can be partitioned into sets $S_0, S_1, \ldots, S_{p-1}$, which we will call cyclic sections, where the probability is 1 of transitioning from a state in one section S_k to a state in the next section S_{k+1} (taking S_p to be S_0), or equivalently if state $i \in S_k$ and $m_{ij} > 0$ imply $j \in S_{k+1}$. In our context, where the circulant matrix A' treats all states equally, all classes must be aperiodic or have the same period.

The averaging scheme in Problem 1 is an example of an irreducible and aperiodic Markov chain where the recurrent class is the set of all positions around the circle. The same can be said for Problems 2 and 3, when n is odd. Otherwise, Problem 2 is not irreducible. It consists of two recurrent classes (the even positions and the odd positions) each of which is aperiodic. In the even case, Problem 3 is irreducible, with one recurrent class that can be partitioned into two cyclic sections (again the even and odd positions), each with period 2.

The limiting behavior of a general Markov chain with transition matrix M depends on the asymptotic behavior of M^k, which in turn is governed by the dominant eigenvalues of M and their corresponding eigenvectors. The results of Perron and Frobenius for eigenvalues of irreducible matrices with nonnegative and with strictly positive elements, are central to the theory of Markov chains for understanding what to expect in general about the dominant eigenvalues of M. We include their results in the following form in order to round out our discussion.

THEOREM A. (Perron–Frobenius) [4]. *Let M be a matrix that is assumed to be irreducible, i.e., its dependency graph is strongly connected.*

(i) If M has (strictly) positive elements, then its eigenvalues can be ordered in such a way that

$$\lambda_0 > |\lambda_1| \geq |\lambda_2| \geq \cdots,$$

and M has a unique dominant eigenvalue; this eigenvalue is positive and simple.

(ii) *If M has nonnegative elements, then its eigenvalues can be ordered in such a way that*

$$\lambda_0 = |\lambda_1| = \cdots = |\lambda_{p-1}| > |\lambda_p| \geq |\lambda_{p+1}| \geq \cdots,$$

and each of the dominant eigenvalues is simple with λ_0 positive. Furthermore, the quantity p is precisely equal to the period of the dependency graph. If $p = 1$, in particular, then there is unicity of the dominant eigenvalue. If $p \geq 2$, the whole spectrum is invariant under the set of transformations

$$\lambda \mapsto \lambda e^{j(2\pi i/p)}, j = 0, 1, \ldots, p-1.$$

The theorem can be applied to each recurrent class of the dependency digraph. In the special case of the stochastic matrix M, we can easily see independently that $\lambda_0 = 1$ is an eigenvalue, and all eigenvalues of M have magnitude less than or equal to one. Thus, from Theorem A, either $\lambda_0 = 1$ is the unique dominant eigenvalue of M, indicating a limit for $M^k \underline{x}_0$ as $k \to \infty$, or all of the dominant eigenvalues of M are equally distributed roots of unity, suggesting an asymptotically periodic behavior for $M^k \underline{x}_0$. Both can be followed up with tools of linear algebra to flesh out the details. Note however, that if M has zero elements, one still needs additional work to determine which of the two alternatives apply.

We are fortunate that in our context, the eigenvalues of a circulant stochastic matrix are directly available through its auxiliary polynomial. In fact, in the following lemma, we completely characterize the dominant eigenvalues of a circulant stochastic matrix and establish the rotational symmetry of its spectrum, all from the auxiliary polynomial alone, independent of Theorem A.

LEMMA. *Let $A = circ[c_0, c_1, \ldots, c_{n-1}]$ be a circulant stochastic matrix with $L = \{i | c_i > 0\}$, $u = \min L = \min\{i | c_i > 0\}$, $L' = \{i - u | c_i > 0\}$, and $g = \gcd(L')$. Then, A has $p = \gcd(n, g)/\gcd(n, g, u)$ dominant eigenvalues $\lambda = \omega^{nj/p} = e^{j(2\pi i/p)}, j = 0, 1, \ldots, p-1$, each having multiplicity* $\gcd(n, g, u)$. *Moreover, the spectrum of A is invariant under the rotational transformation $\lambda \longmapsto \lambda e^{2\pi i/p}$.*

Proof. We generalize a portion of the argument in Section 5 used to solve Problem 3. All eigenvalues of A are of the form $\lambda_j(A) = p_A(\omega^j) = \omega^{uj} \sum_{l \in L'} c_{l+u} \omega^{lj}$ where $c_{l+u} > 0$ for all $l \in L'$. It follows that $\left|p_A(\omega^j)\right| \leq$

1, and $\left|p_A(\omega^j)\right| = 1$ if and only if the second factor $\sum_{l\in L'} c_{l+u}\omega^{lj} = 1$, since both the leading term of the sum being real and $\sum_{l\in L'} c_{l+u} = 1$ together imply that $\sum_{l\in L'} c_{l+u}\omega^{lj} = 1$ (and $\omega^{lj} = 1$ for all $l \in L'$) whenever $\left|\sum_{l\in L'} c_{l+u}\omega^{lj}\right| = 1$. The number of eigenvalues (counting multiplicities) is therefore $\gcd(n, g)$, the number of solutions to $gj \equiv 0 \pmod{n}$ i.e., $j \equiv 0 \pmod{\frac{n}{\gcd(n,g)}}$. (Note: The first nonzero solution and the smallest positive power of ω that yields $\left|p_A(\omega^j)\right| = 1$, is $j = \frac{n}{\gcd(n,g)}$, a fact that will be used in the proof of Theorem 7.)

Now, as j runs through the values $j = \frac{n}{\gcd(n,g)}k$, $k = 0, 1, \ldots, \gcd(n, g) - 1$, the dominant eigenvalues $p_A(\omega^j) = \omega^{uj}$ run through $p = \frac{\gcd(n,g)}{\gcd(n,g,u)}$ distinct values, $\gcd(n, g, u)$ times. Thus, the dominant eigenvalues are the pth roots of unity, each with multiplicity $\gcd(n, g, u)$.

The rotational symmetry of the spectrum is obtained by noting that $\lambda_{j+k}(A) = \omega^{uj+uk} \sum_{l\in L'} c_{l+u}\omega^{lj+lk} = \lambda_j(A)\, e^{2\pi i/p}$, where $p = \frac{\gcd(n,g)}{\gcd(n,g,u)}$, if we can solve simultaneously both $uk \equiv \frac{n}{p} \pmod{n}$ and $gk \equiv 0 \pmod{n}$ for k. But, the second congruence implies $k = \frac{n}{\gcd(n,g)}s$ and the first becomes $\frac{un}{\gcd(n,g)}s \equiv \frac{n\gcd(n,g,u)}{\gcd(n,g)} \pmod{n}$, which has a solution if $\gcd\left(n, \frac{un}{\gcd(n,g)}\right) \Big| \frac{n\gcd(n,g,u)}{\gcd(n,g)}$. In fact, $\gcd\left(n, \frac{un}{\gcd(n,g)}\right) = \gcd\left(\frac{n}{\gcd(n,g)}\gcd(n, g), \frac{n}{\gcd(n,g)}u\right) = \frac{n}{\gcd(n,g)}\gcd(n, g, u)$. Thus, we have rotational symmetry through multiplication by $e^{2\pi i/p}$. $\blacksquare$

We remark that the dominant eigenvalues of A are completely determined by the indices of the positive coefficients of its auxiliary polynomial. Of course, the structure of the transition digraph associated with A guarantees this fact.

7. Invariant based derivation of structure results for the general case. Suppose we wish to construct a simple and easily motivated proof that under appropriately favorable but general conditions, repeated averaging around the circle will result in a limit with values equal to the average of the coordinates of $\underline{x}_0$ (as in Problem 1). We might separate the goal into two tasks: (1) proving that the components of $\underline{x}_k$ approach each other in value, and (2) determining that there is a common limit and its value. The second suggests establishing an invariant, and even suggests the invariant to be used. It is easy to see that as long as A is stochastic, the average of the components of $\underline{x}_k = A^k\underline{x}_0$ remains the same, and thus, must be the common value of the limit.

The first task suggests searching for a useful pseudo-invariant, whose value may change, but whose limit is zero precisely when there is a common limiting value. The obvious candidates are measurements of the coordinate spread of $\underline{x}_k$, such as its range as defined earlier, which we apply next.

THEOREM 5. *Let A be an $n \times n$ stochastic matrix with strictly positive elements. Then, for any column vector $\underline{x}_0$, $A^k \underline{x}_0$ converges as $k \to \infty$ to a column vector with equal components and A^k itself converges to the matrix $\frac{1}{n}J$.*

Proof. Let M_k, m_k, and $R_k = M_k - m_k$ be the maximum and minimum values and the range of $\underline{x}_k = A^k \underline{x}_0$. Further, let a be the value of a minimal entry of A. First, we note that m_k is nondecreasing, for if $b_{i'}^{(k+1)}$ and $b_{i''}^{(k+1)}$ are maximal and minimal components of $\underline{x}_{k+1}$, then

$$m_{k+1} = b_{i''}^{(k+1)} = \sum_{j=0}^{n-1} a_{i''j} b_j^{(k)} \geq \sum_{j=0}^{n-1} a_{i''j} m_k = m_k.$$

Next, we bound the range. If $b_{j'}^{(k)}$ is a minimal component of $\underline{x}_k$, then

$$\begin{aligned} M_{k+1} &= \sum_{j=0}^{n-1} a_{i'j} b_j^{(k)} \leq \sum_{\substack{j=0 \\ j \neq j'}}^{n-1} a_{i'j} M_k + a_{i'j'} m_k = \\ &= \sum_{j=0}^{n-1} a_{i'j} M_k - a_{i'j'} M_k + a_{i'j'} m_k = M_k - a_{i'j'}(M_k - m_k) = \\ &= (1 - a_{i'j'})(M_k - m_k) + m_k \leq (1-a)(M_k - m_k) + m_{k+1}. \end{aligned}$$

Thus,

$$R_{k+1} = M_{k+1} - m_{k+1} \leq (1-a)(M_k - m_k) = (1-a)R_k \leq (1-a)^{k+1} R_0 \to 0$$

as $k \to \infty$. So, $A^k \underline{x}_0$ approaches the vector $\frac{1}{n}(\underline{x}_0' \underline{e})\underline{e}$ and thus, $A^k \to \frac{1}{n}J$. ■

The above proof is similar to that of [6, pp. 448-9] used to prove "the fundamental limit theorem" for regular Markov chains.

Note that Theorem 5 applies to any irreducible and aperiodic matrix A because it can be shown that for a large enough power and beyond, A^k has all positive terms. For example, in Problem 1 we can apply Theorem 5 to A^n, where $a = 1/2^{n-1}$ is its minimum entry.

Unfortunately, the preceding argument does not directly generalize if A is not irreducible (as in Problem 2, n even) or irreducible but not aperiodic (Problem 3, n even) as long as the average is applied to all coordinates of $\underline{x}_0$. Nor would an argument using variance of the coordinates, unless in the case of Problems 2 we split as in Section 5. Nevertheless, we have included the above argument for its elementary nature and for the sake of completeness. More useful will be pseudo-invariants of the form $\|Lx_k\|$, for carefully chosen matrices L dependent on A, which we introduce in the next theorem.

THEOREM 6. *Assume that the stochastic circulant matrix A is used to form $\underline{x}_k = A^k \underline{x}_0$. Then the sequence $||x_k||, k = 0, 1, 2, \ldots$, is nonincreasing. If $\lambda = 1$ is the only dominant eigenvalue of A, and has multiplicity one, then $\underline{x}_k$ converges to a limit vector with identical coordinates and A is irreducible and aperiodic.*

Proof. We set $l_k = \|x_k\|^2$ and consider $d_k = l_k - l_{k+1} = \underline{x}'_k\,(I - A'A)\,\underline{x}_k$, $k \geq 0$. Now, $M = I - A'A$ is positive semi-definite: $|\lambda_j(A)| \leq 1$ for stochastic A, and thus $\lambda_j(M) = 1 - |\lambda_j(A)|^2 \geq 0$ for circulant M. It follows that $d_k \geq 0$ and $\|x_k\|$ is nonincreasing (and in fact, must converge). (Pseudo-invariants of this kind are called monovariants [11].)

Because M is positive semi-definite and symmetric, we can find a matrix square root L such that $L'L = I - A'A = M$. Then for any such square root, $d_k = \|Lx_k\|^2 \to 0$ since $\sum_{k=0}^{\infty} d_k = l_0$ converges. We will use $\|L\underline{x}_k\|$ as our pseudo-invariant.

Now, suppose $\lambda = 1$ is the dominant eigenvalue of A and has multiplicity one. Then, M has eigenvalue $\lambda = 0$ with multiplicity one, and eigenvector $\frac{1}{\sqrt{n}}\underline{e} = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)$. Because M is real and symmetric, we factor $M = U\Lambda U'$ where U is orthogonal with $\frac{1}{\sqrt{n}}\underline{e}$ as its first column and set $L = \Lambda^{1/2}U'$ to be our square root. The one-dimensional null-space of L is generated by $\underline{e}$, for $L\underline{e} = \Lambda^{1/2}U'\underline{e} = \mathrm{diag}\{0, \lambda_1^{1/2}, \ldots, \lambda_{n-1}^{1/2}\}[\sqrt{n}, 0, \ldots, 0]' = \underline{0}$. Thus, $\|Lx_k\|^2 \to 0$ and the invariance of $\frac{1}{n}(\underline{x}'_k\underline{e})$ for all k, yield $\underline{x}_k \to r\underline{e}$ as $k \to \infty$ with $r = \frac{1}{n}(\underline{x}'_0\underline{e})$. Note that we have many choices for L, e.g., we might just as well have taken the symmetric $L = U\Lambda^{1/2}U'$, for $L^2 = L'L = U\Lambda U' = M$. ∎

On the other hand, zero might be a multiple eigenvalue of L if the conditions are relaxed and then $\|L\underline{x}_k\| \to 0$ does not imply the convergence of $\underline{x}_k$ though $\|\underline{x}_k\|$ still converges. Nevertheless, $\|L\underline{x}_k\| \to 0$ may give us valuable information. For example, the special case $A = circ[c_0 = 1/2, 0, \ldots, 0, c_t = 1/2, 0, \ldots, 0]$ generalizes both Problems 1 and 2. For L, we take the circulant matrix $C = I - A = circ[c_0 = 1/2, 0, \ldots, 0, c_t = -1/2, 0, \ldots, 0]$ which corresponds to the pseudo-invariant

$\|C\underline{x}\|^2 = \sum_{i=0}^{n-1} \frac{|x_i - x_{i+t}|^2}{4}$ where the indices are taken mod n. Thus, $\|C\underline{x}_k\|^2 \to 0$ implies asymptotic equality of all the coordinates of $\underline{x}_k$ when $t = 1$ (Problem 1) and when $t = 2$, n odd (half of Problem 2), while in the case $t = 2$ and n even (the other half of Problem 2), the coordinates of $\underline{x}_k$ are partitioned into two sets with asymptotic equality on each set.

The next theorem shows that, in some sense, every averaging scheme around the circle can be reduced to the above special case.

THEOREM 7. *For any circulant stochastic matrix A and any initial configuration $\underline{x}_0$, let u and g be defined as above (in the Lemma). Partition the n positions around the circle into $\gcd(n, g)$ subsets (or one subset, if $\gcd(n, g) = 1$): $S_j = \{s : s \equiv j \pmod{\gcd(n, g)}, 0 \le s < n\}$, $j = 0, 1, \ldots, \gcd(n, g) - 1$, and define the range of each $\underline{x}_k = A^k \underline{x}_0$ when restricted to the subset S_j to be $R_j^{(k)} = \max\left\{b_i^{(k)} : i \in S_j\right\} - \min\left\{b_i^{(k)} : i \in S_j\right\}$. Then, for each j, $R_j^{(k)} \to 0$ as $k \to \infty$.*

Proof. Define t to be the maximum positive value that makes $\left|p_A(\omega^{n/t})\right| = 1$. Then, by the note in the proof of the Lemma, $t = \gcd(n,\ g)$.

We set $M = I - A'A$, as in the proof of Theorem 6, and follow a similar argument except that we choose $C = circ[c_0 = \alpha, 0, \ldots, 0, c_t = -\alpha, 0 \ldots, 0], \alpha > 0$, so that $M = C'C + D$ with D being a positive semi-definite matrix. In fact, we want that $p_D(\omega^k) \ge 0$, for all $k, 0 \le k \le n-1$, with $p_D(x) = p_M(x) - p_{C'C}(x) = p_M(x) - \alpha^2(1 - x^t)(1 - x^{n-t})$. Since $p_M(\omega^k) = 1 - p_A(\omega^k)p_A(\omega^{-k})$, we have $p_M(\omega^k) > 0$ except if $(n/t)|k$ by the Lemma, i.e., when $p_{C'C}(\omega^k) = 0$. Therefore, we can choose α to be the positive square root of

$$\min_{\substack{k:0 \le k \le n-1 \\ \frac{n}{t} \nmid k}} \frac{p_M(\omega^k)}{2 - \omega^{tk} - \omega^{-tk}}.$$

Note that the denominators are positive and real. This will guarantee that $||C\underline{x}_k|| \to 0$ as $k \to \infty$ with

$$||C\underline{x}||^2 = \alpha^2 \sum_{i=0}^{n-1} |x_i - x_{i+t}|^2.$$

The conclusion immediately follows. ∎

We have extracted all that we can from our pseudo-invariant. No doubt, the observant reader has already concluded that the sets S_j are either recurrent classes, or cyclic sections of recurrent classes. But which? We know that all of the numbers at positions in S_j asymptotically behave the same, but we don't know whether they all reach a limit or are asymptotically periodic. As in the proof of Theorem 6, it is when the invariance of the averages (on the S_js in this case) is coupled with the results of the pseudo-invariant that the asymptotic behavior on each subset S_j is revealed, and each can be identified as a recurrent class or a cyclic section. This is the task of the next theorem.

THEOREM 8. *For any circulant stochastic matrix A and any initial configuration $\underline{x}_0$, let u and g be defined as above. Then, the Markov chain with transition matrix A consists of $\gcd(n, g, u)$ recurrent classes, each with period $p = \frac{\gcd(n,g)}{\gcd(n,g,u)}$. In other words, the n positions around the circle can be partitioned into $\gcd(n, g, u)$ rotationally symmetric subsets where, on each subset either the coordinates of $\underline{x}_k$ converge (if $p = 1$) or asymptotically cycle through values with (possibly non-fundamental) period p.*

Proof. The key to the proof is to recognize that in applying the averaging scheme for the $k+1$th time, the coordinates of $\underline{x}_{k+1}$ in positions from S_j are calculated using values from $\underline{x}_k$ in positions from S_{j+u} exclusively (where the subscripts of S are taken mod $\gcd(n, g)$). For instance, if $i \in S_j$ then $i \equiv j \pmod{\gcd(n, g)}$. We calculate $b_i^{(k+1)} = \sum_{l \in L'} c_{l+u} b_{i+l+u}^{(k)}$ and note that for each $l \in L'$, the subscript $i + l + u \equiv j + u \pmod{\gcd(n, g)}$, so position $i + l + u \in S_{j+u}$.

Now, if $u \equiv 0 \pmod{\gcd(n, g)}$ then $S_j = S_{j+u}$, and all of the $S_j s$ are recurrent classes. The average of the numbers at positions from S_j are preserved after each multiplication by A, so by the previous theorem, $\underline{x}_k$ converges and on each recurrent class S_j, its limiting value is the average of the coordinates of $\underline{x}_0$ from S_j. (Note that in this case, $p = \frac{\gcd(n,g)}{\gcd(n,g,u)} = 1$.)

On the other hand, if $u \not\equiv 0 \pmod{\gcd(n, g)}$ then $S_j \neq S_{j+u}$ and it will have taken $p = \frac{\gcd(n,g)}{\gcd(n,g,u)} > 1$ multiplications by A to get $S_j = S_{j+pu}$ since $\frac{\gcd(n,g)}{\gcd(n,g,u)}$ is the smallest positive value of p satisfying $pu \equiv 0 \pmod{\gcd(n, g)}$. Thus, each of the S_js is a cyclic section, with p cyclic sections in each recurrent class, and $\frac{\gcd(n,g)}{p} = \gcd(n, g, u)$ recurrent classes in all. Furthermore, the average of the coordinates of $\underline{x}_{k+1}$ in positions from S_j equals the average of the coordinates of $\underline{x}_k$ in positions from S_{j+u}. Thus, again from the previous theorem $\underline{x}_{kp}$ converges as $k \to \infty$

to identical values at positions from each S_j equal to the average of the coordinates of $\underline{x}_0$ at those same positions. ∎

From Theorem 8 we get a simple asymptotic characterization of the powers of A.

THEOREM 9. *For any circulant stochastic matrix A and any initial configuration $\underline{x}_0$, let u and g be defined as above. Then, through a permutation matrix P, determined by A, the powers of A asymptotically take on the simple form $PA^kP' \approx \frac{\gcd(n,g)}{n} I_{\gcd(n,g,u)} \otimes E_p^k \otimes J_{n/p}$ as $k \to \infty$, where the subscripts indicate the dimensions of the matrices. In terms of the entries of A^k,*

$$a_{ij}^{(k)} \approx \begin{cases} \frac{\gcd(n,g)}{n}, & \textit{if } j - i \equiv ku \pmod{\gcd(n,g)}, \\ 0, & \textit{otherwise}, \end{cases}$$

as $k \to \infty$.

Proof. The proof of this theorem amounts to no more than reinterpreting in matrix form the content of Theorem 8, by reordering of the rows and columns of A according to recurrent classes and cyclic sections within each class. First, for $j = 0, 1, \ldots, t-1$, $t = \gcd(n,g)$, order each cyclic section $S_j = \left(j, j+t, \ldots, j + \left(\frac{n}{t} - 1\right)t\right)$, in increasing order. Next, for $i = 0, 1, \ldots, \gcd(n,g,u) - 1$, order each recurrent class $R_i = S_i \hat{} S_{i+u} \hat{} \ldots \hat{} S_{i+(p-1)u}$ by concatenating its cyclic sections in the order listed, noting that multiplication by A takes values from the positions in each section to give the values in the positions from the immediately preceding section (and from S_i to $S_{i+(p-1)u}$). Then, the concatenation of the recurrent classes $R_0 \hat{} R_1 \hat{} \cdots \hat{} R_{\gcd(n,g,u)-1}$ is a reordering of the numbers 0 through $n-1$.

Finally, we choose the permutation matrix P that makes $P(0, 1, \ldots, n-1)' = (R_0 \hat{} R_1 \hat{} \cdots \hat{}\ R_{\gcd(n,g,u)-1})'$. Then, $PA^kP' \approx \frac{\gcd(n,g)}{n} I_{\gcd(n,g,u)} \otimes E^k \otimes J$ as $k \to \infty$, where I is $\gcd(n,g,u) \times \gcd(n,g,u)$ due to the number of recurrent classes, E is $p \times p$ from the period of each class, and J is $\frac{n}{p} \times \frac{n}{p}$ from the number of states in each cyclic section.

We coax out the limiting values of $a_{ij}^{(k)}$ by noting that the ith component of $\underline{x}_k = A^k(0, \ldots, 0, b_j = 1, 0, \ldots, 0)'$ is nonzero, if and only if $i \in S_{j-ku}$ (where the subscript is taken $\bmod \gcd(n,g)$), and approaches $\gcd(n,g)/n$ (one averaged over the $n/\gcd(n,g)$ positions from S_{j-ku}), as $k \to \infty$. ∎

The last portion of the theorem generalizes Theorem 2 of [8] to include the periodic case.

REMARK 1. Here are some notable examples and features of Theorem 8. If $g = 1$ then A is irreducible and aperiodic (cf. Problem 1 with $u = 0, g = 1$) resulting in a limit vector with equal coordinates. In Problem 3, we get $u = 1,\ g = n - 2$ which yields $\gcd(n, g, u) = 1$ limiting cycle of period $p = 2$ if n is even and $p = 1$ otherwise. If $u = 0$ or $u = g > 0$ then $p = 1$, i.e., A is aperiodic, and thus we have $\gcd(n, g, u) = \gcd(n, g)$ class specific limit vectors, each vector having equal coordinates. If n is even (or odd) and $g = 2$, for instance if L is $\{2, 4, \ldots\}, \{0, 2, 4, \ldots\}$, or $\{0, 2\}$ (cf. Problem 2), then we have two (or one) limit vectors. For example, in Problem 2 with n even, we get $u = 0, g = 2, t = 2, \gcd(n, g, u) = 2$ classes with period $p = \gcd(n, g)/\gcd(n, g, u) = 1$. Thus, we have two separate classes and $\underline{x}_k$ converges to $(E, O, E, O, \ldots, E, O)$. Note that a shift of $u, 0 < u < \gcd(n, g)$ will make room for several dominant eigenvalues (of absolute value one). If $u = 0$ then $\gcd(n, g, u) > 1$ has the same effect. In summary, the multiplicity of eigenvalue $\lambda = 1$ is the number of recurrent classes of the Markov chain, and the number of the unit magnitude eigenvalues, evenly distributed around the unit circle, is the period of each recurrent class (as is true of any recurrent class (with the possibility of different periods) of any general Markov chain).

REMARK 2. Theorem 8 shows that any averaging scheme involving at least two consecutive terms will have the same limit as the limit in Problem 1.

8. Other averaging schemes – general problem. While we have been able to characterize the limiting behavior of any averaging scheme around the circle, with regard to fractionizing, the averaging scheme explored in Section 4 seems to us to be unusually amenable to thorough analysis. In this section, we only touch upon the problem of first fractionizing for two other averaging schemes as appetizers and heartily invite the further exploration of these and any other averaging scheme that might strike the reader's fancy.

Equal weights

Let us take $A = circ[c_0 = 1/N, \ldots, c_{N-1} = 1/N, 0, \ldots, 0]$ with auxiliary polynomial $p_A(x) = \frac{1}{N}(1 + x + \ldots + x^{N-1})$ and

$$p_{A^k}(x) \equiv (p_A(x))^k \equiv \frac{1}{N^k}\left(\frac{1 - x^N}{1 - x}\right)^k \bmod (x^n - 1).$$

To obtain, say, $b_0^{(k)}$ we take the coefficient of x^j in $p_{A^k}(x)$, and get that the relative contribution of $b_j, 0 \leq j \leq \min\{k, n-1\}$,

$$\frac{1}{N^k}\left(\sum_{t\equiv j \bmod n}\sum_{i=0}^{k}(-1)^i\binom{k}{i}\binom{k-1+t-iN}{t-iN}\right). \tag{8.1}$$

This simplifies to $\frac{1}{N^k}\sum_{t\equiv j \bmod n}\binom{k}{t}$ with $N=2$ in Problem 1 (see identity (2.1)). The case with $N>2$ seems a lot more involved than that of $N=2$ because the number of binomial terms in (8.1) increased from one to $k+1$. The problem with only two but not necessarily equal weights seems more manageable.

Two unequal weights

Here we discuss only the following innocent looking case. Let n be an odd prime and set $q = \min_{0\leq j\leq n-1}\rho_n(b_j)$ where $\rho_n(a)$ denotes the highest power of n dividing the integer a. If $A = circ[c_0 = (n-1)/n, c_1 = 1/n, 0, \ldots, 0]$ with $p_A(x) = \frac{1}{n}(n-1+x)$ then

$$p_{A^k}(x) \equiv (p_A(x))^k \equiv \frac{1}{n^k}\sum_{t=0}^{k}\binom{k}{t}(n-1)^{k-t}x^t \bmod (x^n-1).$$

This yields $b_l^{(k)} = \sum_{j=0}^{\min\{k,n-1\}}\sum_{t\equiv j \bmod n}\binom{k}{t}(n-1)^{k-t}b_{l+j}/n^k$ and thus for all $k\leq q$, with some effort

(8.2)

$$b_l^{(k)} \equiv \sum_{j=0}^{\min\{k,n-1\}}(n-1)^{k-j}(-1)^j\left(\sum_{t\equiv j \bmod n}\binom{k}{t}(-1)^t\right)\frac{b_{l+j}}{n^k} \pmod{n^2},$$

by binomial expansion.

We can apply Theorem 3 of [5], i.e., for all $j, 0 \leq j \leq n-1$, $\rho_n(\sum_{t\equiv j \bmod n}\binom{k}{t}(-1)^t) \geq \lfloor\frac{k-1}{n-1}\rfloor$ to increase $\rho_n(b_l^{(k)})$ by increasing k in (8.2). To get a lower bound on the number of steps it takes before fractionizing we can choose the maximum k so that $0 \leq q-k+\lfloor\frac{k-1}{n-1}\rfloor$ which yields the lower bound $q+\lfloor\frac{q-1}{n-2}\rfloor$ for the number of steps it takes before fractionizing.

ACKNOWLEDGMENT. The second author wishes to express his gratitude to the Computer and Automation Research Institute of the Hungarian Academy of Sciences for its hospitality during this work.

References

1. D. S. Clark, *A combinatorial theorem on circulant matrices*, Amer. Math. Monthly **92** (1985), 725–729.

2. L. Comtet, *Advanced Combinatorics*, D. Reidel, Dordrecht, 1974.

3. P. J. Davis, *Circulant Matrices*, Wiley Interscience, New York, 1979.

4. P. Flajolet and R. Sedgewick, *Analytic Combinatorics: Functional Equations*, Rational and Algebraic Functions, INRIA, No. 4103 (January 2001).

5. I. M. Gessel and T. Lengyel, *On the order of Stirling numbers and alternating binomial coefficient sums*, The Fibonacci Quarterly **39** (2001), 444–454.

6. C. M. Grinstead and J. L. Snell, *Introduction to Probability*, 2nd revised edition, American Mathematical Society, Providence, Rhode Island, 1997.

7. G. Iba and J. Tanton, *Candy sharing*, Amer. Math. Monthly **110** (2003), 25–35.

8. O. Krafft and M. Schaefer, *Convergence of the powers of a circulant stochastic matrix*, Linear Algebra Appl. **127** (1990), 59–69.

9. S. Róka, *1500 Elementary Problems in Mathematics*, Typotex, Budapest, 1996. (in Hungarian)

10. G. Tollisen and T. Lengyel, *A congruential identity and the 2-adic order of lacunary sums of binomial coefficients*, INTEGERS, Electronic Journal of Combinatorial Number Theory **A4:4** (2004), 1–8.

11. P. Zeitz, *The art and craft of problem solving*, Wiley, 1999.

Occidental College, Mathematics Department, 1600 Campus Road, Los Angeles, CA 90041
E-mail address: `tollisen@oxy.edu`

Occidental College, Mathematics Department, 1600 Campus Road, Los Angeles, CA 90041
E-mail address: `lengyel@oxy.edu`