Equidistribution of Matrix-Power Residues Modulo One

By Joel N. Franklin

1. Introduction. In numerical analysis artificial random numbers are generated by recurrence formulas of the type

$$x_{n+1} = \{Nx_n + \theta\} \qquad (n = 0, 1, 2, \cdots). \tag{1}$$

Here $\{y\} = y - [y] =$ the fractional part of y. The number N is an integer >1. The number x_0 is a given initial value such that $0 \leqq x_0 < 1$. The number θ is fixed. Some early references to numerical work with sequences of the type (1) are given by O. Taussky and J. Todd in [1]. Regarding the sequence x_n as a function of x_0, I proved in [2] that for almost all x_0 the sequence x_n is equidistributed modulo 1, i.e.,

$$\lim_{k\to\infty} \frac{1}{k} \sum_{a \leqq x_n < b;\, n=0,\cdots,k-1} 1 = b - a \tag{2}$$

whenever $0 \leqq a < b \leqq 1$.

The purpose of this paper is to generalize the preceding result to vector-matrix recurrence formulas

$$x^{(n+1)} = \{Ax^{(n)} + b\} \qquad (n = 0, 1, \cdots). \tag{3}$$

Here each $x^{(n)}$ is a d-dimensional column vector, b is a d-dimensional column vector, and A is a $d \times d$ matrix with integer components. In the preceding case (1), $d = 1$, $A = N$, and $b = \theta$. By $\{y\}$ for a vector y with real components y_i is meant the vector with components $\{y_i\}$. The vector x^0—with parentheses removed around the superscript—is given in the unit cube C_d of d dimensions,

$$C_d{:}0 \leqq x_i < 1 \qquad (i = 1, \cdots, d). \tag{4}$$

All the vectors x^n lie in C_d. The main result of the paper is: *A sufficient condition that x^n be equidistributed for almost all x^0 is that the matrix A be nonsingular and have no eigenvalue which is a root of unity; if $b = 0$, so that $x^{n+1} = \{Ax^n\}$, the condition is necessary as well as sufficient.*

This result has applications to numerical analysis and to the theory of numbers. In [3] the one-dimensional sequences (1) were analyzed at length. It was shown there that for $d > 1$ the successive d-tuples

$$(x_0, \cdots, x_{d-1}), \quad (x_d, \cdots, x_{2d-1}), \quad (x_{2d}, \cdots, x_{3d-1}), \quad \cdots \tag{5}$$

cannot be equidistributed in C_d. In other words, the proportion of these vectors, taken sequentially, which lie in a subregion R of C_d cannot generally be expected to approach the ratio (volume of R)/(volume of C_d) = volume of R. However, as the result stated in the last paragraph shows, if $A = \mathrm{diag}(N, N, \cdots, N)$, where $N =$ integer >1, the vectors defined by (3) are equidistributed for almost all

Received December 5, 1963.

choices of the d components of the initial vector x^0. For example, if $d = 3$ and $b = 0$, we find that the vectors $x^n = (u_n, v_n, w_n)$ $(n = 0, 1, \cdots)$ defined by

$$(6) \qquad u_{n+1} = \{Nu_n\}, \quad v_n = \{Nv_{n+1}\}, \quad w_n = \{Nw_{n+1}\}$$

are equidistributed in the unit cube C_3 for almost all initial values u_0, v_0, w_0.

In the theory of numbers we obtain the following sort of result: For almost all real initial values f_0, f_1, the Fibonacci sequence defined by

$$(7) \qquad f_{n+1} = f_n + f_{n-1} \qquad (n = 1, 2, \cdots)$$

is equidistributed *by twos* modulo one, i.e.,

$$(8) \qquad \lim_{k\to\infty} \frac{1}{k} \sum_{a_1 \leqq f_n < b_1; a_2 \leqq f_{n+1} < b_2; n=0,\cdots,k-1} 1 = (b_1 - a_1)(b_2 - a_2)$$

whenever $0 \leqq a_1 < b_1 \leqq 1$ and $0 \leqq a_2 < b_2 \leqq 1$. Setting $a_2 = 0$, $b_2 = 1$, we obtain the weaker result that almost all Fibonacci sequences are equidistributed modulo one.

2. The Theorems of Weyl and Riesz. A sequence of d-dimensional, real vectors

$$(1) \qquad x^{(n)} = (x_1^n, x_2^n, \cdots, x_d^n) \qquad (n = 0, 1, \cdots)$$

is said to be equidistributed modulo one if

$$(2) \qquad \lim_{k\to\infty} \frac{1}{k} \sum_{a_i \leqq \{x_i^n\} < b_i \quad (i=1,\cdots,d); n=0,\cdots,k-1} 1 = \prod_{i=1}^{d} (b_i - a_i)$$

whenever $0 \leqq a_i < b_i \leqq 1$ $(i = 1, \cdots, d)$. We shall use the following theorem of H. Weyl [4]:

THEOREM. *A sequence* (1) *of d-dimensional vectors $x^{(n)}$ is equidistributed modulo one if and only if*

$$(3) \qquad \lim_{k\to\infty} \frac{1}{k} \sum_{n=0}^{k-1} \exp 2\pi i(j_1 x_1^n + j_2 x_2^n + \cdots + j_d x_d^n) = 0$$

for all combinations of integers $j_1, \cdots, j_d$ except $j_1 = \cdots = j_d = 0$.

We shall also need the ergodic theorem of F. Riesz; see [5] and [2]:

THEOREM. *Let a measurable set Ω be given, of finite or infinite measure, the corresponding measure and integral being defined according to Lebesgue, or more generally, by means of a distribution of positive masses. That being the case, let us designate by T a point-transformation which is single-valued (but not necessarily one-to-one) from Ω onto itself; and let us suppose that T conserves measure in the sense that, E being a measurable set, TE its transform, and E' the set of points P whose images appear in TE, the sets E' and TE have the same measure. Then, if $f_1(P)$ is an integrable function and $f_k(P) = f_1(T^{k-1}P)$, the arithmetic mean of the functions $f_1, f_2, \cdots, f_n$ converges almost everywhere, as $n \to \infty$, to an integrable function $\phi(P)$ which is invariant (almost everywhere) under T. If Ω is of finite measure,*

$$(4) \qquad \int_\Omega \phi(P) = \int_\Omega f_1(P).$$

3. Measure-Preserving Congruences Modulo One. Let A be a $d \times d$ matrix with real components, and let b be a d-component column vector. We define a transformation $y = Tx$ of the d-dimensional unit cube C_d into itself by the congruence

$$y \equiv Ax + b \pmod 1 \tag{1}$$

by which we mean $y = \{Ax + b\}$ or, equivalently,

$$y_i \equiv \sum_{j=1}^{d} a_{ij}x_j + b_i \pmod 1 \qquad (i = 1, \cdots, d).$$

We wish to determine when this transformation is measure-preserving.

First we remark that the congruence (1) is measure-preserving if and only if the congruence

$$w \equiv Ax \pmod 1 \tag{2}$$

is measure-preserving. That is because the congruence (1) may be composed of two transformations, $w = \{Ax\}$ and $y = \{w + b\}$. Since the second transformation is one-to-one and measure-preserving, the composite transformation (1) is measure-preserving if and only if the first transformation (2) is measure-preserving.

Second, we remark that the transformation T is measure-preserving if and only if

$$\int_{C_d} f(P) = \int_{C_d} f(TP) \tag{3}$$

for all scalar functions f which are measurable in C_d. This elementary remark is justified by Riesz in [5].

LEMMA. *Let K be the set of nonzero d-dimensional column-vectors k with integer components. Let K_1 be the set of d-dimensional real column vectors with at least one component equal to a nonzero integer. Then the congruence $y \equiv Ax + b \pmod 1$ is measure-preserving in C_d if and only if the transpose matrix A^* maps K into K_1.*

Proof. Let the measurable function $f(P) = f(x)$ have the Fourier series

$$f(x) \sim c(0) + \sum_{k \in K} c(k) \exp 2\pi i k^* x. \tag{4}$$

Since the Fourier series is multiply periodic, the congruence T is measure-preserving if and only if

$$c(0) = \int_{C_d} f(x)\, dx = \int_{C_d} f(Ax)\, dx \tag{5}$$

for all measurable f. But

$$\begin{aligned} \int_{C_d} f(Ax)\, dx &= c(0) + \sum_{k \in K} c(k) \int_{C_d} \exp 2\pi i k^* A x\, dx \\ &= c(0) + \sum_{k \in K} c(k) \int_{C_d} \exp 2\pi i (A^* k)^* x\, dx. \end{aligned} \tag{6}$$

Therefore, T is measure-preserving if and only if

$$\int_{C_d} \exp 2\pi i(A^*k)^*x \, dx = 0 \qquad \text{for all} \quad k \in K \tag{7}$$

which is true if and only if $A^*k \in K_1$ for all $k \in K$.

The lemma shows that, if $d = 1$, the congruence $y \equiv Ax + b$ is measure-preserving if and only if A is a nonzero integer. However, if $d > 1$, the matrix A may have noninteger coefficients. For example, the congruence

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \equiv \begin{pmatrix} 0 & -6 \\ \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \pmod 1$$

is measure-preserving. To see this, we observe that

$$A^*k = \begin{pmatrix} \frac{1}{2}k_2 \\ -6k_1 + k_2 \end{pmatrix}.$$

If $k \in K$, the first component $k_2/2$ is a nonzero integer unless k_2 is zero or odd. If $k_2 = 0$, the second component $= -6k_1 =$ integer $\neq 0$; if k_2 is odd, $-6k_1 + k_2 =$ even integer $+$ odd integer $\neq 0$. Therefore, A^* maps K into K_1.

In the rest of the paper we shall suppose that A has all components equal to integers.

THEOREM. *If all the components of A are integers, the congruence $y \equiv Ax + b$ (mod 1) is measure-preserving if and only if* $\det A \neq 0$.

Proof. This result follows immediately from the lemma. Since A has integer components, if $\det A = 0$ there is a vector $k \in K$ such that $A^*k = 0$, which is not in K_1. If $\det A \neq 0$, all vectors A^*k are nonzero vectors with integer components when $k \in K$, so that $A^*k \in K \in K_1$.

4. Ergodic Congruences Modulo One. We shall say that a measure-preserving transformation $y = Tx$ from the d-dimensional unit cube into itself is *ergodic* if the only measurable functions $\phi(x)$ for which

$$\phi(x) = \phi(Tx) \quad \text{almost everywhere in } C_d \tag{1}$$

are the functions $\phi(x) =$ constant a.e. (almost everywhere).

LEMMA. *Let B be a $d \times d$ matrix with integer components. Let K be the set of nonzero d-dimensional column-vectors with integer components. Then the sequence of vectors k, Bk, B^2k, $\cdots$ is unbounded for every k in K if and only if B has no eigenvalue which is zero or a root of unity.*

Proof. Suppose that for some k in K the sequence B^jk is bounded. Since B and k have integer components, each of the vectors B^jk must be one of the finite number of integer-component vectors which lie in some bounded subset of d-dimensional Euclidean space. Therefore, $B^rk = B^sk$ for some $r > s$. If B has no zero eigenvalue, B is nonsingular and $B^qk = k$ for $q = r - s$. But then

$$0 = \det (B^q - I) = \prod_{j=0}^{q-1} \det (B - \omega^j I)$$

where $\omega = \exp (2\pi i/q)$. Then one of the roots of unity ω^j is an eigenvalue of B.

Conversely, if B has a zero eigenvalue, since B has integer components, there is an eigenvector k in K such that $0 = Bk = B^2k = \cdots$, a bounded sequence. If B has an eigenvalue which is a qth root of unity, then B^q has 1 as an eigenvalue. Then there is an eigenvector k in K such that $B^qk = k$, and the sequence B^jk is periodic, hence bounded.

THEOREM. *Let A be a nonsingular $d \times d$ matrix with integer components, and let b be a d-dimensional column-vector with real components. Then the measure-preserving congruence $y \equiv Ax + b \pmod 1$ is ergodic if A has no eigenvalue which is a root of unity. The congruence $y \equiv Ax \pmod 1$ is ergodic if and* only *if A has no eigenvalue which is a root of unity.*

Proof. Let $Tx \equiv Ax + b \pmod 1$, where b is a vector with real components, and A is a nonsingular matrix with integer components and with no eigenvalue equal to a root of unity. Then $B =$ transpose of $A = A^*$ has no eigenvalue which is zero or a root of unity. According to the lemma, B^jk is unbounded as $j \to \infty$ for every k in K.

Let $\phi(x)$ be any measurable function satisfying (1). Since T is measure-preserving,

$$\phi(x) = \phi(T^jx) \text{ a.e. for all } j = 1, 2, \cdots. \tag{2}$$

The measurable function $\phi(x)$ has a Fourier series

$$\phi(x) \sim a(0) + \sum_{k \in K} a(k) \exp 2\pi i k^* x. \tag{3}$$

Furthermore,

$$T^jx \equiv A^jx + b^{(j)} \pmod 1 \tag{4}$$

where $b^{(j)} = b + Ab + \cdots + A^{j-1}b$. Therefore,

$$\phi(T^jx) \sim a(0) + \sum_{k \in K} a(k) \exp 2\pi i k^*(A^jx + b^{(j)})$$

or, equivalently, with $B = A^*$,

$$\phi(T^jx) \sim a(0) + \sum_{k \in K} (a(k) \exp 2\pi i k^* b^{(j)}) \exp 2\pi i (B^jk)^* x. \tag{5}$$

Therefore,

$$\begin{aligned} a(k) \exp 2\pi i k^* b^{(j)} &= \int_{C_d} \phi(T^jx) \exp(-2\pi i (B^jk)^* x)\, dx \\ &= \int_{C_d} \phi(x) \exp(-2\pi i (B^jk)^* x)\, dx. \end{aligned} \tag{6}$$

Since B^jk is unbounded for each k in K, the integrals (6) tend to zero for some subsequence of j tending to ∞. But the left-hand side of (6) has modulus $|a(k)|$ for all j. Therefore, $a(k) = 0$ for all $k \in K$. Then the Fourier series for $\phi(x)$ consists only of the constant term $a(0)$. Therefore, $\phi(x)$ equals this constant almost everywhere.

If $Tx \equiv Ax \pmod 1$, i.e., if $b = 0$, we can show that the transformation is ergodic *only* if A has no eigenvalue which is a root of unity. Suppose that A, and therefore B, have eigenvalues which are qth roots of unity. Then $B^qk = k$ for some

k in K. Let p be the smallest positive integer such that $B^p k = k$. Since A, and therefore B, is nonsingular, no two of the vectors, $k, Bk, \cdots, B^{p-1}k$ are equal. Therefore, the function

$$\phi(x) = \sum_{j=0}^{p-1} \exp\,(2\pi i k^* A^j x) \tag{7}$$

is nonconstant. But $\phi(x) = \phi(Tx)$, since $k^*A^p = (B^p k)^* = k^*$. Therefore, T is not ergodic. This completes the proof of the theorem.

If $b \neq 0$, the transformation $Tx \equiv Ax + b \pmod 1$ may be ergodic even if A has an eigenvalue which is a root of unity. For example, the transformation $Tx \equiv x + b$ is ergodic if and only if the components of b are rationally independent, i.e., if $k^*b \neq$ integer for all k in K. This result follows immediately from the uniqueness of the Fourier series of a measurable function $\phi(x)$.

A more interesting question arises when $A \neq I$. For example, consider the transformation

$$Tx \equiv \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ \sqrt{2} \end{pmatrix} \pmod 1. \tag{8}$$

If $\phi(x)$ has the Fourier series (3), then

$$\phi(T^j x) \sim a(0) + \sum_{k \in K} a_j(k) \exp 2\pi i (2^j k_1 x_1 + k_2 x_2)$$

where $a_j(k) = a(k) \exp 2\pi i k_2 \sqrt{2}$. Then the invariance (1) implies

$$a_j(k) = \int_0^1 \int_0^1 \phi(x) \exp -2\pi i (2^j k_1\, x_1 + k_2\, x_2)\, dx_1\, dx_2 \,.$$

Letting $j \to \infty$, we see that $a(k) = 0$ unless $k_1 = 0$. But then

$$\phi(x_1, x_2) \sim \sum_{k_2 \neq 0}^{\infty} a(0, k_2) \exp 2\pi i k_2 x_2 \,.$$

Now the irrationality of $\sqrt{2}$ implies that $a(0, k_2) = 0$ for all $k_2 \neq 0$. Therefore, the transformation (8) is ergodic.

THEOREM. *Let*

$$\begin{aligned} y_1 &\equiv N x_1 + b_1 \pmod 1 \\ y_s &\equiv x_s + b_s \qquad (s = 2, \cdots, d) \end{aligned} \tag{9}$$

where N is an integer with absolute value >1, and the b_s are real. This measure-preserving transformation is ergodic if and only if $k_2 b_2 + \cdots + k_d b_d \neq$ integer for any integers $k_2, \cdots, k_d$ which are not all zero.

Proof. This theorem is an immediate and obvious generalization of the preceding example.

5. Equidistribution of Matrix-Power Residues.

THEOREM. *Let A be a $d \times d$ matrix with integer components. Let b be a d-dimensional column vector with real components. Given the vector $x = x^{(0)}$, construct the sequence $x^{(j)}$ by the recurrence formula*

(1) $$x^{(j+1)} \equiv Ax^{(j)} + b \pmod 1$$

for $j = 0, 1, \cdots$. *This sequence is equidistributed modulo one for almost all* x *if* A *has no eigenvalue equal to zero or a root of unity; if* $b = 0$, *the sequence is equidistributed for almost all* x *if and* only *if* A *has no eigenvalue equal to zero or a root of unity.*

Proof. If A has no eigenvalue equal to zero, A is nonsingular; and, according to the theorem in Section 3, the transformation $Tx \equiv Ax + b \pmod 1$ is measure-preserving. Therefore, by the Riesz ergodic theorem, for all measurable functions f

(2) $$\frac{1}{k}\sum_{j=0}^{k-1} f(x^{(j)}) \to \phi(x) \quad \text{as} \quad k \to \infty$$

for almost all $x = x^{(0)}$, where $\phi(x) = \phi(Tx)$ a.e. By the first theorem in Section 4, if A is nonsingular and has no eigenvalue which is a root of unity, $\phi(x) =$ constant a.e. By the Riesz ergodic theorem, since the d-dimensional unit cube C_d has finite measure $= 1$, the constant ϕ has the integral

(3) $$\int_{C_d} f(x)\, dx = \int_{C_d} \phi\, dx = \phi.$$

If $0 \leqq a_i < b_i \leqq 1$ $(i = 1, \cdots, d)$ define

(4) $$\begin{aligned} f(x) = f(x_1, \cdots, x_d) &= 1 \qquad \text{for } a_i \leqq x_i < b_i \qquad (i = 1, \cdots, d) \\ &= 0 \qquad \text{elsewhere in } C_d. \end{aligned}$$

From (2) and (3) we have the result, for almost all x, that the sequence $x^{(j)}$ is equidistributed in C_d.

For $b = 0$ we must prove the "only if" part of the theorem. First suppose that A has an eigenvalue equal to zero. Then $A^*k = 0$ for some k in K. Let

(5) $$f(x) = \exp 2\pi i k^* x.$$

Since $f(x)$ is Riemann-integrable, we must have

(6) $$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} f(x^{(j)}) = \int_{C_d} f(x)\, dx$$

if $x^{(j)}$ is equidistributed; for a proof of this result see Koksma [6]. From (5) we have

(7) $$f(x^{(j)}) = \exp 2\pi i k^* A^j x = 1 \qquad (j \geqq 1).$$

Therefore, the limit on the left-hand side of (6) equals one. Since the integral of $f(x)$ equals zero, equation (6) is false; and the sequence $x^{(j)}$ cannot be equidistributed.

Finally, for $b = 0$ suppose that A is nonsingular but that A has an eigenvalue which is a root of unity. Construct the nonconstant, Riemann-integrable function $\phi(x)$ defined in formula (7) of Section 4. Since $\phi(x) = \phi(Tx)$, we have

(8) $$\frac{1}{n}\sum_{j=0}^{n-1} \phi(x^{(j)}) = \phi(x^{(0)}) = \phi(x) \qquad \text{for all } n.$$

But

(9) $$\int_{C_d} \phi(x)\, dx = 0.$$

Therefore, the sequence $x^{(j)}$ cannot be equidistributed. This completes the proof of the theorem.

6. Application to Numerical Analysis. In Monte Carlo calculations in d dimensions, the basic property required of pseudo-random vectors $x^{(j)}$ is usually the property (6) of Section 5. This property is equivalent to the equidistribution of the $x^{(j)}$. The reader is now referred back to the next to the last paragraph of Section 1.

7. Equidistribution of Fibonacci Sequences. We shall say that a sequence of real numbers x_n is *equidistributed by d's* modulo one if the sequence of successive d-tuples

$$x^{(n)} = \begin{pmatrix} x_{n+1} \\ x_{n+2} \\ \vdots \\ x_{n+d} \end{pmatrix} \qquad (n = 0, 1, \cdots) \tag{1}$$

is equidistributed modulo one, as defined in Section 2. This concept was considered at length in [3]. For $d = 1$ we have the usual definition for the equidistribution of x_n modulo one. A sequence equidistributed by d's for $d > 1$ is equidistributed by r's for $1 \leqq r < d$, but the converse is false.

THEOREM. *Let a general Fibonacci sequence x_n be defined by*

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_d x_{n-d} \qquad (n > d) \tag{2}$$

where $a_1, a_2, \cdots, a_d$ are integers. Then for almost all real initial values $x_1, \cdots, x_d$ the sequence x_n is equidistributed by d's modulo one if and only if

$$z^d \neq a_1 z^{d-1} + a_2 z^{d-2} + \cdots + a_d \tag{3}$$

for $z = 0$ or for $z =$ a root of unity.

Proof. Define the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ a_d & a_{d-1} & a_{d-2} & \cdots & a_1 \end{bmatrix}. \tag{4}$$

The relation (2) is equivalent to the vector-matrix relation

$$x^{(n+1)} = A x^{(n)} \qquad (n = 0, 1, \cdots). \tag{5}$$

The eigenvalues of A are the roots of the equation

$$0 = \det (zI - A) = z^d - a_1 z^{d-1} - \cdots - a_d . \tag{6}$$

The theorem now follows directly from the result in Section 5.

California Institute of Technology
Pasadena, California

1. O. Taussky & J. Todd, "Generation of pseudo-random numbers," *Symposium on Monte Carlo Methods*, H. A. Meyer, Editor, John Wiley and Sons, New York, 1956, p. 15–18.
2. J. N. Franklin, "On the equidistribution of pseudo-random numbers," *Quart. Appl. Math.*, v. 16, 1958, p. 183–188.
→ J. N. Franklin, "Deterministic simulation of random processes," *Math. Comp.*, v. 17, 1963, p. 28–59.
→ H. Weyl, "Über die Gleichverteilung von Zahlen modulo Eins," *Math. Ann.*, v. 77, 1916, p. 313–352.
5. F. Riesz, "Sur la théorie ergodique," *Comment. Math. Helv.*, v. 17, 1945, p. 221.
6. J. F. Koksma, *Diophantische Approximationen*, Chelsea, New York, 1936.