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

\[ x_{n+1} = \lfloor Nx_n + \theta \rfloor \quad (n = 0, 1, 2, \ldots). \]

Here \( \lfloor y \rfloor = y - \lfloor y \rfloor \) is the fractional part of \( y \). The number \( N \) is an integer \( > 1 \). The number \( x_0 \) is a given initial value such that \( 0 \leq x_0 < 1 \). The number \( \theta \) 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.,

\[ \lim_{k \to \infty} \frac{1}{k} \sum_{a \leq x_n < b, n = 0, \ldots, k-1} 1 = b - a \]

whenever \( 0 \leq a < b \leq 1 \).

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

\[ x^{(n+1)} = \lfloor Ax^{(n)} + b \rfloor \quad (n = 0, 1, \ldots). \]

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 \( \lfloor y \rfloor \) for a vector \( y \) with real components \( y_i \), is meant the vector with components \( \lfloor y_i \rfloor \). The vector \( x^{(0)} \)—with parentheses removed around the superscript—is given in the unit cube \( C_d \) of \( d \) dimensions,

\[ C_d: 0 \leq x_i < 1 \quad (i = 1, \ldots, d). \]

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 \( x \) is that the matrix \( A \) be nonsingular and have no eigenvalue which is a root of unity; if \( b = 0 \), so that \( x^{n+1} = \lfloor Ax^n \rfloor \), 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

\[ (x_0, \ldots, x_{d-1}), \quad (x_d, \ldots, x_{2d-1}), \quad (x_{2d}, \ldots, x_{3d-1}), \quad \ldots \]

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 \( (\text{volume of } R)/(\text{volume of } C_d) \) = volume of \( R \). However, as the result stated in the last paragraph shows, if \( A = \text{diag}(N, N, \ldots, N) \), where \( N = \text{integer } > 1 \), 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

$$u_{n+1} = \lfloor Nu_n \rfloor, \quad v_n = \lfloor Nv_{n+1} \rfloor, \quad w_n = \lfloor Nw_{n+1} \rfloor$$

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

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

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

$$\lim_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} 1 = (b_1 - a_1)(b_2 - a_2)$$

whenever $0 \leq a_1 < b_1 \leq 1$ and $0 \leq a_2 < b_2 \leq 1$. 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 $(1) x^{(n)} = (x_1^n, x_2^n, \cdots, x_d^n)$ ($n = 0, 1, \cdots$) is said to be equidistributed modulo one if

$$\lim_{k \to \infty} \frac{1}{k} \sum_{a_1 \leq x_1^n < b_1, \cdots, a_d \leq x_d^n < b_d} 1 = \prod_{i=1}^{d} (b_i - a_i)$$

whenever $0 \leq a_i < b_i \leq 1$ ($i = 1, \cdots, d$). 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

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

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 $\Omega$ 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 $\Omega$ 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 $\Omega$ is of finite measure,

$$\int_\Omega \phi(P) = \int_\Omega f_1(P).$$
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
\[ y = Ax + b \pmod{1} \]
by which we mean $y = \{Ax + b\}$ or, equivalently,
\[ y_i = \sum_{j=1}^{d} a_{ij}x_j + b_i \pmod{1} \quad (i = 1, \cdots, d). \]
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
\[ w = Ax \pmod{1} \]
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
\[ \int_{C_d} f(P) = \int_{C_d} f(TP) \]
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 $y = Ax + b \pmod{1}$ 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
\[ f(x) \sim c(0) + \sum_{k \in K} c(k) \exp 2\pi ik^*x. \]
Since the Fourier series is multiply periodic, the congruence $T$ is measure-preserving if and only if
\[ c(0) = \int_{C_d} f(x) \, dx = \int_{C_d} f(Ax) \, dx \]
for all measurable $f$. But
\[ \int_{C_d} f(Ax) \, dx = c(0) + \sum_{k \in K} c(k) \int_{C_d} \exp 2\pi ik^*Ax \, dx \]
\[ = c(0) + \sum_{k \in K} c(k) \int_{C_d} \exp 2\pi i(A^*k)^*x \, dx. \]
Therefore, $T$ is measure-preserving if and only if
(7) \( \int_{C_d} \exp 2\pi i(A^*k)^*x \, dx = 0 \) for all \( k \in K \)

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 \( y = Ax + b \) 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

\[
\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 & -6 \\ \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad \text{(mod 1)}
\]

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

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

are the functions \( \phi(x) = \text{constant a.e. (almost everywhere)} \).

**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 \( k, Bk, B^2k, \ldots \) 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 \( B^r k \) is bounded. Since \( B \) and \( k \) have integer components, each of the vectors \( B^r k \) must be one of the finite number of integer-component vectors which lie in some bounded subset of \( d \)-dimensional Euclidean space. Therefore, \( B^r k = B^s k \) for some \( r > s \). If \( B \) has no zero eigenvalue, \( B \) is nonsingular and \( B^q k = k \) 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 \( \omega^j \) 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 $q$th 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 $B^j k$ 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 $y = Ax + b \pmod{1}$ is ergodic if $A$ has no eigenvalue which is a root of unity. The congruence $y = Ax \pmod{1}$ is ergodic if and only if $A$ has no eigenvalue which is a root of unity.

**Proof.** Let $Tx = Ax + b \pmod{1}$, 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 $B = \text{transpose of } A = A^*$ has no eigenvalue which is zero or a root of unity. According to the lemma, $B^j k$ 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,

\[(2) \quad \phi(x) = \phi(T^j x) \text{ a.e. for all } j = 1, 2, \ldots .\]

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

\[(3) \quad \phi(x) \sim a(0) + \sum_{k \in K} a(k) \exp 2\pi ik^*x .\]

Furthermore,

\[(4) \quad T^j x = A^j x + b^{(j)} \pmod{1}\]

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

\[\phi(T^j x) \sim a(0) + \sum_{k \in K} a(k) \exp 2\pi ik^*(A^j x + b^{(j)})\]

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

\[(5) \quad \phi(T^j x) \sim a(0) + \sum_{k \in K} (a(k) \exp 2\pi ik^*b^{(j)}) \exp 2\pi i(B^j k)^*x .\]

Therefore,

\[a(k) \exp 2\pi ik^*b^{(j)} = \int_{c^d} \phi(T^j x) \exp (-2\pi i(B^j k)^*x) \, dx\]

\[= \int_{c^d} \phi(x) \exp (-2\pi i(B^j k)^*x) \, dx .\]

Since $B^j k$ is unbounded for each $k$ in $K$, the integrals (6) tend to zero for some subsequence of $j$ tending to $\infty$. 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 $Tx = Ax \pmod{1}$, 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 $q$th roots of unity. Then $B^q k = k$ 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

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

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 \( Tx = Ax + b \mod 1 \) may be ergodic even if \( A \) has an eigenvalue which is a root of unity. For example, the transformation \( Tx = x + b \) 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

$$T_x = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ \sqrt{2} \end{pmatrix} \mod 1.$$  

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

$$\phi(x_1, x_2) \sim \sum_{k_2 \neq 0} a(0, k_2) \exp 2\pi i k_3 x_2.$$  

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

$$y_1 \equiv Nx_1 + b_1 \mod 1$$
$$y_s \equiv x_s + b_s, \quad (s = 2, \cdots, d)$$

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
\[
\begin{align*}
\text{(1) } & \quad x^{(j+1)} \equiv Ax^{(j)} + b \pmod{1} \\
\text{for } j = 0, 1, \ldots. \text{ This sequence is equidistributed modulo one for almost all } x \text{ if } A \\
& \text{has no eigenvalue equal to zero or a root of unity; if } b = 0, \text{ the sequence is equidistributed} \\
& \text{for almost all } x \text{ if and only if } A \text{ has no eigenvalue equal to zero or a root of unity.}
\end{align*}
\]

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

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

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, \( \phi(x) = \text{constant} \) a.e. By the Riesz ergodic theorem, since the \( d \)-dimensional unit cube \( C_d \) has finite measure = 1, the constant \( \phi \) has the integral

\[
\int_{C_d} f(x) \, dx = \int_{C_d} \phi \, dx = \phi.
\]

If \( 0 \leq a_i < b_i \leq 1 \) \((i = 1, \ldots, d)\) define

\[
\begin{align*}
&\begin{align*}
\phi(x) = f(x_1, \ldots, x_d) = 1 & \quad \text{for } a_i \leq x_i < b_i \quad (i = 1, \ldots, d) \\
&= 0 & \quad \text{elsewhere in } C_d.
\end{align*}
\end{align*}
\]

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

\[
\phi(x) = \exp 2\pi ik^* x.
\]

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

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

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

\[
\int_{C_d} f(x^{(j)}) \, dx = \exp 2\pi ik^* A^j x = 1 \quad (j \geq 1).
\]

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

\[
\frac{1}{n} \sum_{j=0}^{n-1} \phi(x^{(j)}) = \phi(x^{(0)}) = \phi(x) \quad \text{for all } n.
\]

But

\[
\int_{C_d} \phi(x) \, dx = 0.
\]
Therefore, the sequence \( x^{(f)} \) 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 \( x^{(f)} \) is usually the property (6) of Section 5. This property is equivalent to the equidistribution of the \( x^{(f)} \). 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

\[
(1) \quad x^{(n)} = \begin{pmatrix} x_{n+1} \\ x_{n+2} \\ \vdots \\ x_{n+d} \end{pmatrix} \quad (n = 0, 1, \ldots)
\]

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 \( 1 \leq r < d \), but the converse is false.

**Theorem.** Let a general Fibonacci sequence \( x_n \) be defined by

\[
(2) \quad x_n = a_1x_{n-1} + a_2x_{n-2} + \cdots + a_dx_{n-d} \quad (n > d)
\]

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

\[
(3) \quad z^d \neq a_1z^{d-1} + a_2z^{d-2} + \cdots + a_d
\]

for \( z = 0 \) or for \( z = \) a root of unity.

**Proof.** Define the matrix

\[
(4) \quad A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_d & a_{d-1} & a_{d-2} & \cdots & a_1 \end{pmatrix}
\]

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

\[
(5) \quad x^{(n+1)} = Ax^{(n)} \quad (n = 0, 1, \ldots).
\]

The eigenvalues of \( A \) are the roots of the equation

\[
(6) \quad 0 = \det (zI - A) = z^d - a_1z^{d-1} - \cdots - a_d.
\]

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

California Institute of Technology
Pasadena, California