Mixing properties of triangular feedback shift registers

Bernd Schomburg*

Abstract

The purpose of this note is to show that Markov chains induced by non-singular triangular feedback shift registers and non-degenerate sources are rapidly mixing. The results may directly be applied to the post-processing of random generators and to stream ciphers in CFB mode.

1 Introduction

Let $\mathcal{S} = (K_t)_{t\geq 0}$ be a (memoryless) source, i.e. a sequence of identically distributed, independent random variables. Informally, a source is an object that emits symbols (in general from a finite alphabet) according to some random mechanism. This could be, for example, a physical random generator or - in first-order approximation - a natural alphabet-based (plain) text. The objective of this paper is to investigate the Markov behaviour of a class of feedback shift registers when the input symbols are modified by the random source. To model the influence of the random source we will assume that the underlying alphabet carries a group structure. We will show that for non-degenerate sources and non-singular triangular feedback shift registers, the associated Markov chains are rapidly mixing. We will also give estimates for mixing (convergence) rate using the Dobrushin coefficients of the associated stochastic matrices.

2 Feedback shift registers and associated Markov chains

In this section we introduce the class of triangular feedback shift registers and derive some of their basic algebraic properties. Furthermore we will define Markov chains associated with feedback shift registers.

*Capucienenlaan 25, 1950 Kraainem, Belgium, e-mail: bernd.schomburg@tiscali.be

Let Σ be a finite alphabet. A **feeedback shift register (FSR)** of length N over Σ is a finite automaton (X, ϕ, λ) with state space $X = \Sigma^N$, a **transition function** $\phi : X \to X$ which is of the form

$$\phi = s \circ \phi_0, \ \phi_0 : X \to X$$

(with the cyclic shift $s : X \to X, \ (x_1, ..., x_N) = (x', x_N) \mapsto (x_N, x_1, ..., x_{N-1}) = (x_N, x')$) and the **output function**

$$\lambda = pr_1 \circ \phi = pr_N \circ \phi_0 : \ X \to \Sigma.$$

Definition 2.1. Let $\mathcal{R} = (X, \phi, \lambda)$ be an FSR.
(i) $\mathcal{R}$ is called **non-singular** iff ϕ (or equivalently ϕ_0) is bijective.
(ii) An FSR (X, ϕ, λ) is called **triangular** iff ϕ_0 has the form

$$\phi_0(x_1, ..., x_N) = (f_1(x_1), f_2(x_1, x_2), ..., f_N(x_1, ..., x_N)). \tag{2.1}$$

with mappings $f_j : \Sigma^j \to \Sigma, \ j = 1, ..., N$.

Remark 2.2. For a family $(f_j : \Sigma^j \to \Sigma)_{1 \le j \le N}$ define $F_j : \Sigma^j \to \Sigma^j$, $j \in \{1, ..., N\}$, by

$$F_j(x_1, ..., x_j) = (f_1(x_1), f_2(x_1, x_2), ..., f_j(x_1, ..., x_j)).$$

If F_k is bijective, then all $F_j, j < k$, are bijective, too. To simplify notation, we will write f for f_N and F for F_{N-1}.

Remark 2.3. (i) Let $(\Sigma, *)$ be a group, $h_j : \Sigma^{j-1} \to \Sigma, \quad j \ge 2$ mappings and $h_1 \in \Sigma$. The family $(f_j)_{1 \le j \le N}$, defined by

$$f_j(x_1, ..., x_j) = x_j * h_j(x_1, ..., x_{j-1}) \quad \forall (x_1, ..., x_j) \in \Sigma^j, 1 \le j \le N,$$

induces a non-singular triangular FSR via (2.1).
(ii) For $\Sigma = \mathbb{F}_2$ the converse is also true: For a non-singular triangular shift register the functions f_j are necessarily of the form

$$f_j(x_1, ..., x_j) = x_j + h_j(x_1, ..., x_{j-1}),$$

where $+$ denotes the addition (modulo 2) in $\mathbb{F}_2$.

Let (X, ϕ, λ) be an FSR. We will investigate the state sequence $\xi = (\xi_j)_{j \ge 0} \in X^{\mathbb{N}_0}$

$$\xi_j = \phi^j x, \quad j \ge 0, \tag{2.2}$$

(belonging to the initial state $x \in X$) and the corresponding output sequence $\alpha = (\alpha_j)_{j \ge 0} \in \Sigma^{\mathbb{N}_0}$

$$\alpha_j = \lambda(\xi_j) = (\lambda \circ \phi^j) \ x, \quad j \ge 0. \tag{2.3}$$

To this end we introduce the mappings

$$s_a : X \to X, (x_1, ..., x_N) \mapsto (a, x_1, ..., x_{N-1}), \quad a \in \Sigma. \tag{2.4}$$

The following lemma will be our basic tool.

Lemma 2.4. Let (X, ϕ, λ) be triangular. For $x \in X$ let ξ and α be defined as in (2.2) and (2.3), respectively. If $y \in X$ an arbitrary state and if we define inductively

$$\eta_0 = y, \quad \eta_{j+1} = s_{\alpha_j} \circ \phi_0(\eta_j), \ j \geq 0, \tag{2.5}$$

then for each $j \in \{1, ..., N\}$ the first j components of ξ_j and η_j coincide

$$pr_k(\xi_j) = pr_k(\eta_j) \quad \forall k \in \{1, ..., j\}.$$

Especially we have

$$\phi^N x = \xi_N = \eta_N = ((s_{\alpha_{N-1}} \circ \phi_0) \circ ... \circ (s_{\alpha_0} \circ \phi_0))y. \tag{2.6}$$

Proof (by induction w.r.t. j).
(a) $j = 1$. By (2.2), (2.3) and (2.4) we have $pr_1(\xi_1) = \alpha_0 = pr_1(\eta_1)$.

(b) $j \to j + 1$. Let $j \in \{1, ..., N - 1\}$ and

$$pr_k(\xi_j) = pr_k(\eta_j) \qquad \forall k \in \{1, ..., j\}.$$

By (2.1) we have

$$pr_k(\phi_0 \xi_j) = pr_k(\phi_0 \eta_j) \quad \forall k \in \{1, ..., j\}$$

and thus

$$pr_k(\xi_{j+1}) = pr_k(\eta_{j+1}) \quad \forall k \in \{2, ..., j + 1\}.$$

Finally, Definition (2.3) of α_j implies

$$pr_1(\xi_{j+1}) = \alpha_j = pr_1(\eta_{j+1})$$

and the assertion follows. □

Remark 2.5. As a first application of Lemma 2.4 we show that for a non-singular triangular FSR $\mathcal{R} = (X, \phi, \lambda)$ and an initial state $x \in X$ the sequences ξ and α have the same period. Recall that for a sequence $g = (g_i)_{i \in \mathbb{N}_0}$ in a set D its **period** $per(g) \in \mathbb{N} \cup \{\infty\}$ is defined by

$$per(g) = \inf\{l \in \mathbb{N} \mid \forall i \in \mathbb{N}_0 : \ g_i = g_{i+l}\}.$$

If the period $per(g)$ is finite, it divides each l with $g_i = g_{i+l} \quad \forall i \in \mathbb{N}_0$. Since $\mathcal{R}$ is non-singular $per(\xi)$ is equal to the (finite) length of the orbit $x^{<\phi>}$ (thus dividing the order of ϕ). Furthermore, trivially

$$l := per(\alpha) \mid per(\xi).$$

(2.6) now implies

$$\phi^{N+l} x = \phi^N x$$

and therefore (using the bijectivity of ϕ) $\phi^l x = x$ and $\xi_{j+l} = \xi_j \quad \forall j \geq 0$, i.e. $per(\xi) \mid l = per(\alpha)$, which finally shows that $per(\xi) = per(\alpha)$.

For the remaining part of the paper we assume that $(\Sigma, *)$ is a **group**. Our plan is to define stochastic processes associated with FSRs. First fix some notation. For a (non-empty) finite set D let $\mathcal{M}_1(D)$ denote the set of all measures on D with mass 1 (also referred to as (probability) distributions on D). The **uniform distribution** on D will be denoted by γ_D, i.e. $\gamma_D(\{\omega\}) = \frac{1}{|D|}$ for all $\omega \in D$.

Definition and Remark 2.6 We define stochastic processes associated with an FSR $\mathcal{R} = (X, \phi, \lambda)$ of length N over Σ. Let $\mathcal{S} = (K_t)_{t\geq 0}$ be a memoryless source, i.e. a sequence of independent, identically distributed (i.i.d.) random variables over Σ with (common) distribution $\sigma \in \mathcal{M}_1(\Sigma)$. Fix an arbitrary random variable Z_0 on X and let the stochastic process $\mathcal{Z} = (Z_t)_{t\geq 0}$ be recursively defined by

$$Z_t = (S_{K_{t-1}} \circ \phi) Z_{t-1}, \quad t \geq 1, \tag{2.7}$$

where $S_a : X \to X$, $a \in \Sigma$ is given by

$$S_a(x_1, ..., x_n) = (a * x_1, ..., x_n), \forall (x_1, ..., x_n) \in X. \tag{2.8}$$

By construction $\mathcal{Z}$ is a homogenous Markov chain over X with transition matrix $P = P(\phi, \sigma) = (P_{xy})_{x,y\in X}$:

$$P_{xy} = Pr(Z_t = y \mid Z_{t-1} = x) = \begin{cases} \sigma(a), & if\ (S_a \circ \phi)(x) = y, \\ 0, & else. \end{cases} \tag{2.9}$$

We will call $\mathcal{Z}$ the **Markov chain associated with $\mathcal{R}$, $\mathcal{S}$ and Z_0**. Note that $P \in \mathbb{R}^{X\times X}$ is stochastic (cf. Definition 3.1) and depends on $\mathcal{R}$ and σ, only. Furthermore, if $\mathcal{R}$ is non-singular, P is doubly-stochastic (cf. Definition 3.2 (i)); indeed, since ϕ is bijective, we have

$$\sum_{x\in X} P_{xy} = \sum_{\bar{x}\in X} \left\{ \begin{array}{ll} \sigma(a), & if\ S_a\bar{x} = y, \\ 0, & else. \end{array} \right\} = \sum_{a\in\Sigma} \sigma(a) = 1$$

for all $y \in X$.
If μ is the distribution of Z_0, μP^j is the distribution of Z_j.
In view of (2.9), mappings of the form

$$\Phi_a = (S_{a_{j-1}} \circ \phi) \circ ... \circ (S_{a_0} \circ \phi), a = (a_0, ...a_{j-1}) \in \Sigma^j, j \in \mathbb{N},$$

will play an important role in the investigation of the powers of P.

Theorem 2.7. Let (X, ϕ, λ) be a triangular FSR and F be defined as in Remark 2.2.
(i) For bijective F, the semi-group $\langle S_a \circ \phi \mid a \in \Sigma\rangle$, generated by all compositions $S_a \circ \phi$, acts transitively on X; more precisely, for each pair $x, y \in X$ there exists a uniquely determined $a = (a_0, ..., a_{N-1}) \in \Sigma^N = X$ with

$$x = \Phi_a y = ((S_{a_{N-1}} \circ \phi) \circ ... \circ (S_{a_0} \circ \phi))y. \tag{2.10}$$

(ii) If $y \in X$ and $1 \leq j \leq N$, then

$$\Sigma^j \ni a \mapsto \Phi_a\ y \in X$$

is one-to-one.
(iii) If F is not bijective, then for all $x \in \Sigma \times (\Sigma^{N-1} \setminus F(\Sigma^{N-1}))$, $y \in X$ and $a \in \Sigma^j$, $j \in \mathbb{N}$:

$$x \neq \Phi_a y.$$

Proof. (i) Let $\psi, \psi_0 : X \to X$ denote the bijections which are defined by $(x', x_N) \mapsto (x_N, Fx')$ and $(x', x_N) \mapsto (Fx', x_N)$, respectively, and consider the corresponding register (X, ψ, pr_N). Let $x, y \in X$ and $(\alpha_j)_{j\geq 0}$ be the output sequence, which corresponds to the initial state $\psi^{-N}x$:

$$\alpha_j = pr_N(\psi^{j-N}x). \tag{2.11}$$

Define (η_j) recursively by $\eta_0 = y$ and

$$\eta_{j+1} = (S_{a_j} \circ \phi)(\eta_j),\ a_j = \alpha_j * f(\eta_j)^{-1}, j \geq 0. \tag{2.12}$$

Then

$$\eta_{j+1} = (a_j * f(\eta_j), F\eta_j') = (\alpha_j, F\eta_j') = (s_{\alpha_j} \circ \psi_0)\eta_j.$$

Lemma 2.4, applied to (X, ψ, pr_N), shows

$$x = \psi^N(\psi^{-N}x) = \eta_N = ((S_{a_{N-1}} \circ \phi) \circ ... \circ (S_{a_0} \circ \phi))y.$$

In particular, we have shown that the mapping

$$X = \Sigma^N \ni a \mapsto \Phi_a y \in X$$

is onto. Since X is finite, the mapping is also one-to-one, which proves that representation in (2.10) is unique.
(ii) Set $A_j = \{\Phi_a\ y \mid a \in \Sigma^j\}$ and note that the assertion is equivalent to

$$|A_j| = |\Sigma|^j. \tag{2.13}$$

For the proof of (2.13) we use induction w.r.t j:
For $j = N$, (2.13) is a consequence of (2.10). Now assume that (2.13) is true for a $j > 1$. Then $|A_{j-1}| \leq |\Sigma|^{j-1}$. Furthermore,

$$\Sigma \times A_{j-1} \ni (a, x) \mapsto (a * f(x), Fx') \in A_j$$

is onto, thus

$$|\Sigma| \cdot |A_{j-1}| \geq |A_j| = |\Sigma|^j,$$

i.e.$|A_{j-1}| \geq |\Sigma|^{j-1}$.
(iii) Clear from the definition of Φ_a. □

Remark 2.8. (i) The group structure $*$ on Σ induces a natural group structure $\hat{*}$ on $X = \Sigma^N$. If $\phi : X \to X$ is a group homomorphism w.r.t. $\hat{*}$, then there is a group homomorphism $\Psi : X \to X$ s.t.

$$\Phi_a x = \Psi(a)\hat{*}\phi^N x \quad \forall a, x \in X.$$

For bijective F the homomorphism Ψ is bijective by Theorem 2.7(i).
(ii) Let F be bijective. Theorem 2.7 (ii) shows that for all $j \leq N$

$$(P^j)_{xy} = \begin{cases} \prod_{i<j} \sigma(a_i), & if\ \Phi_a(x) = y, (a_0, ..., a_{j-1}) \in \Sigma^j \\ 0, & else. \end{cases}$$

Especially, if we define the product distribution $\hat{\sigma} \in \mathcal{M}_1(X)$ by

$$\hat{\sigma}(a) = \prod_{i=0}^{N-1} \sigma(a_i) \quad \forall a = (a_0, ..., a_{N-1}) \in X, \tag{2.14}$$

then

$$(P^N)_{xy} = \hat{\sigma}(a) \; if \; \Phi_a(x) = y, a \in X. \tag{2.15}$$

If $\sigma = \gamma_\Sigma$, P has the limiting distribution (cf. Definition 3.2 (ii)) γ_X with $\mu P^j = \gamma_X$ for all $\mu \in \mathcal{M}_1(X)$ and $j \geq N$.
(iii) For traditional FSRs we have $f_i = pr_i$ for $1 \leq i < N$, i.e. $F = id_{\Sigma^{N-1}}$, so the prerequisites of Theorem 2.7 (i) are automatically fulfilled. As an application of the proof of 2.7 consider the case that

$$\phi(x', x_N) = (c, x') \;\; \forall (x', x_N) \in X$$

for a constant $c \in \Sigma$. The constuction (2.11)-(2.12) shows that for given x, y the uniquely determined a with $\Phi_a(x) = y$ is given by

$$a_j = y_{N-j} * c^{-1}.$$

Thus P has the limiting distribution π

$$\pi = \hat{\sigma}(\cdot \hat{*} (c, ..., c)^{-1}) \tag{2.16}$$

with $\mu P^j = \pi$ for all $\mu \in \mathcal{M}_1(X), \; j \geq N$. Furthermore $\pi \neq \gamma_X$ if $\sigma \neq \gamma_\Sigma$.

Proposition 2.9. Let F_k be bijective for a $k < N$ and let $y, \bar{y} \in X$ s.t. $y_1 \neq \bar{y}_1$. Then $\Phi_a \; y \neq \Phi_{\bar{a}} \; \bar{y}$ for all $j \in \{1, ..., k\}$ and $a, \bar{a} \in \Sigma^j$.

This follows immediately from Remark 2.2 and the following

Lemma 2.10. Let $k \in \{2, ..., N\}$ s.t F_{k-1} is bijective. If $y, \bar{y} \in X$, $j \in \mathbb{N}$, and $(a_0, ..., a_{j-1}), (\bar{a}_1, ..., \bar{a}_{j-1}) \in \Sigma^j$ with

$$pr_i(\Phi_{(a_0,...,a_{j-1})} y) = pr_i(\Phi_{(\bar{a}_0,...,\bar{a}_{j-1})} \bar{y}) \quad \forall i \in \{1, ..., k\},$$

then

$$pr_i(\Phi_{(a_0,...,a_{j-2})} y) = pr_i(\Phi_{(\bar{a}_0,...,\bar{a}_{j-2})} \bar{y}) \quad \forall i \in \{1, ..., k-1\}.$$

Proof. Let $x = \Phi_{(a_0,...,a_{j-2})} y$ and $\bar{x} = \Phi_{(\bar{a}_0,...,\bar{a}_{j-2})} \bar{y}$. Then

$$\begin{array}{lll} \Phi_{(a_0,...,a_{j-1})} y & = (S_{a_{j-1}} \circ \phi) x & = (a_{j-1} * f_N(x), f_1(x_1), ..., f_{N-1}(x_1, ..., x_{N-1})), \\ \Phi_{(\bar{a}_0,...,\bar{a}_{j-1})} \bar{y} & = (S_{\bar{a}_{j-1}} \circ \phi) \bar{x} & = (\bar{a}_{j-1} * f_N(\bar{x}), f_1(\bar{x}_1), ..., f_{N-1}(\bar{x}_1, ..., \bar{x}_{N-1})). \end{array}$$

By assumption,

$$f_i(x_1, ..., x_i) = f_i(\bar{x}_1, ..., \bar{x}_i) \quad \forall i \in \{1, ..., k-1\}.$$

and F_{k-1} is one-to-one, so the assertion follows. □

3 Stochastic matrices

The purpose of this section is to recall some notions and results from the theory of Markov chains. For the convenience of the reader we include a proof of Theorem 3.5 which is the main result of this section and which will be used in Section 4. For further details and proofs we refer to [1] and [2].

Let $n \in \mathbb{N}$. In the real linear space $\mathbb{R}^n$ $(e_i)_{1\leq i\leq n}$ denotes the canonical basis and $(\epsilon_i)_{1\leq i\leq n}$ the corresponding dual basis. We set $e = \sum_{i=1}^n e_i$. We identify $\mathcal{M}_1^n = \mathcal{M}_1(\{1,...,n\})$ with the set of all $\ell \in (\mathbb{R}^n)^*$ with $\ell \geq 0$[1] and $\ell(e) = 1$ by $\mu \leftrightarrow \sum_{i=1}^n \mu(\{i\})\epsilon_i$. Then $\gamma := \frac{1}{n}\sum_{i=1}^n \epsilon_i$ is the uniform distribution on $\{1,...,n\}$. We endow $\mathcal{M}_1^n$ with the metric

$$d_1(\mu,\nu) = \|\mu - \nu\|_1 = \sum_i |\mu_i - \nu_i|.$$

Since $\mathcal{M}_1^n$ is closed in $(\mathbb{R}^n)^*$, $(\mathcal{M}_1^n, d_1)$ is a complete metrc space.

Definition and Remark 3.1. A matrix $P \in \mathbb{R}^{n\times n}$ is **stochastic**, iff $P \geq 0$ and $Pe = e$, or equivalently, iff $\mu P \in \mathcal{M}_1^n$ for all $\mu \in \mathcal{M}_1^n$. Thus, for a stochastic $P \in \mathbb{R}^{n\times n}$, the operator

$$T_P : \mathcal{M}_1^n \ni \mu \mapsto \mu P \in \mathcal{M}_1^n$$

is well-defined and continuous. It is convenient to define certain properties of stochastic matrices using the language of dynamical systems in metric spaces. Let (X,d) be a metric space and $T : X \to X$ a continuous operator. A point $x \in X$ is called **global attractor** of T iff $x = \lim_{j\to\infty} T^j y$ for each $y \in X$; in this case x is the only fixed-point of T.

Definition 3.2. Let $P \in \mathbb{R}^{n\times n}$ be stochastic.
(i) P is called **doubly-stochastic**, iff γ is a fixed-point of T_P, i.e. iff its transposed matrix ${}^tP$ is stochastic, too. (ii) A distribution $\pi \in \mathcal{M}_1^n$ is called **stationary** w.r.t. P iff π is a fixed-point of T_P. (iii) A distribution $\pi \in \mathcal{M}_1^n$ is called **limiting distribution** of P, iff π is the global attractor of T_P. (iv) P is called **ergodic**, iff it has a limiting distribution > 0.

Remark 3.3. If a stochastic P has a limiting distribution π, then by 3.1 π is the only stationary distribution of P. If, in addition, P is doubly-stochastic, then $\pi = \gamma$ and P is ergodic.

Definition 3.4. Let P be a stochastic $n\times n$-matrix. The **Dobrushin** (or **ergodicity**) **coefficient** is defined to be

$$\delta(P) = \frac{1}{2}\max_{i<j}\sum_{k=1}^n |P_{ik} - P_{jk}| = \max_{i<j}\frac{\|\epsilon_i P - \epsilon_j P\|_1}{\|\epsilon_i - \epsilon_j\|_1}.$$

[1] For a (non-empty) finite set I and $x, y \in \mathbb{R}^I$ we write $x \leq y$ iff $x_i \leq y_i$ $\forall i \in I$, and $x < y$ iff $x_i < y_i$ $\forall i \in I$.

Obviously, $0 \leq \delta(P) \leq 1$, and

$$\delta(P) = 0 \Leftrightarrow \exists \pi \in \mathcal{M}_1^n : \ P = e \otimes \pi, \tag{3.1}$$

$$\delta(P) = 1 \Leftrightarrow \exists i, j : \ i < j \ \wedge \ \min(\epsilon_i P, \epsilon_j P) = 0. \tag{3.2}$$

Furthermore, it can be shown that

$$\|\mu P - \nu P\|_1 \leq \delta(P) \|\mu - \nu\|_1 \ \forall \mu, \nu \in \mathcal{M}_1^n. \tag{3.3}$$

For permutation matrices $A, B \in \mathbb{R}^{n \times n}$ the product BPA is stochastic with

$$\delta(BPA) = \delta(P). \tag{3.4}$$

Theorem 3.5. Let $P \in \mathbb{R}^{n \times n}$ be a stochastic matrix.
(i) P has a limiting distribution iff $\delta(P^N) < 1$ for an $N \in \mathbb{N}$.
(ii) If $\delta(P^N) < 1$ for an $N \in \mathbb{N}$ and if $\pi \in \mathcal{M}_1^n$ is the limiting distribution of P (which exists by (i) and is unique by definition), then

$$\|\mu P^j - \pi\|_1 \leq \delta(P^N)^{\lfloor \frac{j}{N} \rfloor} \|\mu - \pi\|_1 \ \forall \mu \in \mathcal{M}_1^n$$

for all $j \geq 0$.

Proof. (i) "⇒": Assume that $\delta(P^N) = 1$ for all $N \in \mathbb{N}$. Then, by (3.2), for each N there exist indices $i_N < j_N$ with $\min(\epsilon_{i_N} P^N, \epsilon_{j_N} P^N) = 0$. By the pigeon hole principle there are indices $i < j$ and a sequence $(N_k)_{k \geq 1} \subset \mathbb{N}$ s.t. $\lim_{k \to \infty} N_k = \infty$ and

$$\min(\epsilon_i P^{N_k}, \epsilon_j P^{N_k}) = 0. \tag{3.5}$$

Now assume that P has a limiting distribution $\pi \in \mathcal{M}_1^n$. Then $\pi = \lim_{k \to \infty} \epsilon_i P^{N_k} = \lim_{k \to \infty} \epsilon_j P^{N_k}$, so by (3.5) $\pi = 0$, which is a contradiction.
(i) "⇐" and (ii): Set $T = T_P$ and let $N \in \mathbb{N}$ s.t. $L = \delta(P^N) < 1$. Then T^N is contractive with Lipschitz constant L. By Banach's fixed-point theorem, T^N has a global attractor $\pi \in \mathcal{M}_1^n$. Now π is the only fixed-point of T^N and $T\pi = T(T^N \pi) = T^N(T\pi)$, so $T\pi = \pi$. Finally, let $\mu \in \mathcal{M}_1^n$, $j \geq 0$ and write j in the form $j = kN + r$ with $k = \lfloor \frac{j}{N} \rfloor$, $0 \leq r < N$. Then, using (3.3) and $\delta(P) \leq 1$,

$$\begin{aligned} d_1(T^j \mu, \pi) &= d_1(T^j \mu, T^j \pi) = d_1(T^{Nk} T^r \mu, T^{Nk} T^r \pi) \\ &\leq L^k d_1(T^r \mu, T^r \pi) \leq L^k d_1(\mu, \pi). \end{aligned}$$ □

Definition and Remark 3.6. A stochastic matrix P is called **primitive**, iff there there is an N with $P^N > 0$. By definition, every ergodic stochastic matrix is primitive. Conversely, if P is primitive with $P^N > 0$, then $\delta(P^N) < 1$, so P has a limiting distribution π, i.e. $\lim_{j \to \infty} P^j e_k = \pi(e_k)$ for all k. It is now easy to see that

$$\min_i (P^j x)_i \leq \min_i (P^{j+1} x)_i \quad \forall x \in \mathbb{R}^n, j \geq 0,$$

so $0 < \min_i (P^N e_k)_i \leq \pi(e_k) \ \ \forall k \in \{1, ..., n\}$, i.e. $\pi > 0$ and P is ergodic. Note that a stochastic matrix is primitive iff a corresponding Markov chain is irreducible and aperiodic.

4 Mixing properties of FSRs

We are now ready to combine the results of Sections 2 and 3. As in Section 2, let $\mathcal{R} = (X, \phi, \lambda)$ be an FSR of length N over a finite group $(\Sigma, *)$, $\mathcal{S} = (K_t)_{t\geq 0}$ a sequence of i.i.d. random variables over Σ with (common) distribution $\sigma \in \mathcal{M}_1(\Sigma)$ and Z_0 be an arbitrary random variable on X with distribution $\mu \in \mathcal{M}_1(X)$. Consider the Markov chain $\mathcal{Z} = (Z_t)_{t\geq 0}$ associated with $\mathcal{R}$, $\mathcal{S}$ and Z_0 and the corresponding stochastic matrix $P = P(\phi, \sigma)$. We assume that $\mathcal{R}$ is triangular and let F_k, F be defined as in 2.2.

Theorem 4.1. (i) If F_k $(k < N)$ is bijective, then

$$\delta(P^j) = 1 \quad \forall j \in \{0, ..., k\}. \tag{4.1}$$

(ii) For bijective F every row of P^N is a permutation of the tuple

$$\left(\hat{\sigma}(a) \mid a \in \Sigma^N = X\right).$$

(iii) For bijective F and non-degenerate σ (i.e. $\min_a \sigma(a) > 0$) the matrix P is primitive. In fact $P^N > 0$ (and thus $\delta(P^N) < 1$).
(iv) If F is not bijective, P is not primitive.

Proof. (i) Proposition 2.9 shows that for $j < k$ P^j has (at least) two rows with disjoint supports. (4.1) now follows from (3.2).(ii) and (iii) follow from (2.15). (iv) is a consequence of Theorem 2.7 (iii). □

We now come to the announced theorem concerning the rapid mixing of non-singular tringular FSRs under the influence of non-degenerate sources. The theorem actually holds under the slightly milder hypothesis that F is bijective and is a direct consequence of Theorems 3.5 and 4.1 (iii) and Remark 3.6.

Theorem 4.2. Let F be bijective and σ non-degenerate. Then P is ergodic with a limiting distribution $\pi > 0$. If $\mathcal{R}$ is non-singular, $\pi = \gamma_X$. For the distribution μP^j of Z_j we have the estimate

$$\|\mu P^j - \pi\|_1 \leq \|\mu - \pi\|_1 \delta(P^N)^{\lfloor \frac{j}{N} \rfloor} \leq C\ \delta(P^N)^{\lfloor \frac{j}{N} \rfloor} \quad \forall j \geq 0,$$

where

$$C = \begin{cases} 2\frac{|X|-1}{|X|}, & if\ \pi = \gamma_X, \\ 2, & else, \end{cases}$$

and $\delta(P^N) < 1$. □

Obviously, the determination of $\delta(P^N)$ plays a vital role in the analysis of concrete FSRs. In order to shed some light on this practical problem, we conclude this note with some remarks on "linear" and binary FSRs.

Theorem 4.3. Let $\mathcal{R}$ be non-singular and linear in the sense that ϕ is a group isomorphism. Then $P = P(\phi, \sigma)$ fulfills

$$\delta(P^N) = \frac{1}{2} \max_{x \in X} \sum_y |\hat{\sigma}(y) - \hat{\sigma}(y \hat{*} x^{-1})| =: h_N(\Sigma, \sigma). \tag{4.2}$$

Proof. By Remark 2.8 (i) there is a bijection Ψ s.t.

$$\Phi_a x = \Psi(a) \hat{*} \phi^N x \quad \forall a, x \in X.$$

Thus

$$y = \Phi_a x \Leftrightarrow a = \Psi^{-1}(y \hat{*} (\phi^N x)^{-1})$$

and (cf. (2.15))

$$(P^N)_{xy} = \hat{\sigma}(\Psi^{-1}(y \hat{*} (\phi^N x)^{-1})).$$

By (3.4)

$$\delta(P^N) = \delta(Q)$$

where

$$Q_{xy} = \hat{\sigma}(y \hat{*} x^{-1}).$$

The Dobrushin coefficient of Q is now easily calculated as

$$\delta(Q) = \frac{1}{2} \max_{x \in X} \sum_y |\hat{\sigma}(y) - \hat{\sigma}(y \hat{*} x^{-1})|. \qquad \square$$

Remark 4.4. An important special case are binary FSRs. Let $(\Sigma, *) = (\mathbb{F}_2, +)$ and let σ be a non-degenerate distribution on $\mathbb{F}_2$, $p = \max(\sigma(0), \sigma(1)) < 1$ and $q = \min(\sigma(0), \sigma(1)) > 0$. Recall that $\varepsilon(\sigma) := \sigma(0) - \sigma(1)$ is called the bias of σ. Then the maximum in (4.3) is attained for $x = (1, 1, ..., 1)$ with

$$h_N(\mathbb{F}_2, \sigma) = \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \binom{N}{k} (p^{N-k} q^k - p^k q^{N-k}).$$

Note that

$$h_{N+1}(\mathbb{F}_2, \sigma) - h_N(\mathbb{F}_2, \sigma) = \begin{cases} \binom{N}{\frac{N}{2}} (pq)^{\frac{N}{2}} (p - q), & \textit{if } N \textit{ is even}, \\ 0, & \textit{if } N \textit{ is odd}. \end{cases}$$

Therefore $\lim_{N \to \infty} h_N(\mathbb{F}_2, \sigma) = 1$ (use the Taylor expansion of $(1 - 4x)^{-\frac{1}{2}}$ around $x = 0$). By Theorem 4.1(ii)

$$\delta(P(\phi, \sigma)^N) \leq h_N(\mathbb{F}_2, \sigma)$$

for all ϕ with bijective F. Computer experiments suggest the following

Conjecture. *For all non-constant Boolean functions $f : \mathbb{F}_2^N \to \mathbb{F}_2$ we have*

$$\delta(P(\phi_f, \sigma)^N) \geq |\varepsilon(\sigma)|$$

where

$$\phi_f(x) = (f(x), x'), x = (x', x_N) \in \mathbb{F}_2^{N-1} \times \mathbb{F}_2,$$

the lower bound being attained e.g. for $f = \chi_{\{0\}}$, the characteristic function of the singleton $\{0\}$.

Note that we have $\delta(P(\phi_f, \sigma)^N) = 0$ for constant f (cf. (2.16) and (3.1)).

Remark 4.5. The source $\mathcal{S}$ can be interpreted as a random source or (a first-order approximation of) a plain text source. The process $\mathcal{Z}$ is then a model of the sequence of states of the FSR (X, ϕ, λ) when operated in cipher-feedback (CFB) mode. The results of this section may therefore be applied to the post-processing of random generators and to stream ciphers in CFB mode.

References

[1] Behrends, E.: *Introduction to Markov Chains.* Vieweg, Braunschweig/Wiesbaden, 2000.

[2] Brémaud, P.: *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queus.* Springer-Vlg., New York etc., 1999.