Perturbing Subshifts of Finite Type

Nick Ramsey *

DePaul University

Contents

1 Introduction ........................................... 1
2 Parry measure and the weight function \( w(\sigma) \) ........................................... 2
3 Bounding entropy ...................................... 4
4 Blocking ................................................... 5
5 Growing words .......................................... 6

1 Introduction

Let \( n \) be a positive integer and let \( T \) be an irreducible \( n \times n \) matrix with entries in \( \{0, 1\} \). This determines a subshift of finite type \( \Sigma \) defined as the collection of all bi-infinite strings \( (x_i) \) on an alphabet of \( n \) symbols (indexing the rows and columns of \( T \)) that are admissible in the sense that the \( (x_{i+1}, x_i) \) entry in \( T \) is equal to 1 for all \( i \in \mathbb{Z} \). The goal of this note is to explore the entropy \( h(\Sigma) \) of the shift map on \( \Sigma \) and how it is affected by perturbations obtained by forbidding various finite words from occurring in \( \Sigma \). Recall that the largest eigenvalue of \( T \) is given by \( \lambda = e^{h(\Sigma)} \), and we assume throughout that \( \lambda > 1 \).

Let \( S \) denote a finite set of finite admissible nonempty words none of which contains another, and let \( \Sigma(S) \) denote the subshift of \( \Sigma \) consisting of those elements of \( \Sigma \) in which none of the words in \( S \) appear. In [1], Lind addressed the problem of bounding the entropy of \( \Sigma(S) \) in case \( S \) consists of a single word. He proved that the entropy of \( \Sigma(S) \) approaches that of \( \Sigma \) as the length \( \ell \) of the word tends to infinity, and showed moreover that the difference in entropy is of order \( \lambda^{-\ell} \). In [3], the author adapted Lind’s method to the case where \( S \) has more than one word, and in particular introduced a certain determinant of correlation polynomials whose size is closely tied to the entropy perturbation. Analyzing the size of this determinant gets complicated as \( S \) grows, and we were only able to effectively bound the entropy and show that it approximates that of \( \Sigma \) well in case \( S \) consists of two words of length tending to infinity. The entropy perturbation in this case is shown to be of order at most \( \lambda^{-\ell} \) where \( \ell \) is the length of the shorter word.

In [2], Miller introduced a different approach to the following related problem: given a finite set \( S \) of finite nonempty words in an alphabet, determine whether there exists a bi-infinite word in the alphabet that avoids \( S \). Note that the ambient shift here is constrained to the full shift, while the set \( S \) is quite flexible. In this note, we adapt Miller’s method to a general subshift \( \Sigma \) and refine it to get lower bounds on the entropy of the perturbations \( \Sigma(S) \). As in [2], we define

\[
p(t) = \sum_{\tau \in S} t^{\|\tau\|}
\]

Theorem 1. There exists a constant \( C \) depending only on \( \Sigma \) such that if \( k \) is a positive integer, every element of \( S \) has length at least \( k \), and there exists \( t \in (1, \lambda^k) \) with

\[
r = \frac{1 + kC\lambda^{2k}p(1/k/\lambda)}{t} < 1
\]

* nramsey@depaul.edu
2 Parry measure and the weight function \( w(\sigma) \)

Let \( u \) and \( v \) denote left and right \( \lambda \)-eigenvectors for \( T \) normalized so that \( u^t v = 1 \). The entries of these vectors measure the prominence of the corresponding symbols as a sink and source in \( \Sigma \), respectively. More precisely, \( u_i / \sum u_i \) is the fraction of paths on the directed graph associated to \( T \) that terminate at \( i \), while \( v_i / \sum v_i \) is the fraction of paths that begin at \( i \). Let \( \mu \) denote Parry measure on \( \Sigma \). This is a shift-invariant measure of maximal entropy and can be characterized on cylinder sets by

\[
\mu([\sigma]) = \frac{u_i v_j}{\lambda^k}
\]

The notation \([\sigma]\) only defines a cylinder set up to shifts. The shift-independence of \( \mu \) often, as here, renders this ambiguity moot. Where it is important to have an actual set to work with (e.g. in defining \( f_\sigma \) below) we take \( \sigma \) to begin at coordinate 0 in forming \([\sigma]\). If \( \sigma \) fails to be admissible, then \([\sigma]\) is taken to be empty.

Let \( N(\sigma, k) \) denote the number of words \( \eta \) of length \( k \) such that \( \sigma \eta \) is admissible. Since \( T \) is irreducible, there exist positive constants \( A, B \) such that

\[
A \lambda^k \leq N(\sigma, k) \leq B \lambda^k
\]

for all words \( \sigma \) and all \( k \geq 1 \). Let \( D \) denote the maximum ratio among the \( v_i \).

**Lemma 1.** We have

\[
\mu([\sigma \tau]) \leq DA^{-1} \lambda^{-|\tau|} \mu([\sigma])
\]

for all words \( \sigma, \tau \).

**Proof.** The explicit description of \( \mu \) on cylinder sets implies that \( D \) is also the maximum ratio among the \( \mu([\sigma \eta]) \) as \( \eta \) varies among words of a given positive length. Thus

\[
\lambda^{|\tau|} \mu([\sigma \tau]) \leq A^{-1} N(\sigma, |\tau|) \mu([\sigma]) = A^{-1} \sum_{|\eta|=|\tau|} \mu([\sigma \eta]) \]

\[
\leq A^{-1} \sum_{|\eta|=|\tau|} D \mu([\sigma \eta]) = DA^{-1} \mu([\sigma])
\]

Let \( \sigma \) be an admissible word and let \([\sigma]\) denote the associated cylinder set. Define a polynomial-valued function \( f_\sigma \) on \([\sigma]\) by

\[
f_\sigma(\alpha) = \sum_{\tau \in S} \sum j \lambda^j
\]

where the inner sum is over \( j \geq 0 \) such that \( \tau \) occurs in \( \alpha \) beginning within \( \sigma \) and ending \( j \) symbols beyond the end of \( \sigma \). Observe that the function \( f_\sigma \) is locally constant on \([\sigma]\).
We define a weight function on admissible words by

\[ w(\sigma) = \frac{1}{\mu([\sigma])} \int_{[\sigma]} f_\sigma \]

Note that the empty word \( \sigma_0 \) has cylinder set \( [\sigma_0] = \Sigma \) and weight 0. Since no element of \( S \) contains another, for each \( j \geq 0 \) there is at most one element of \( S \) that ends \( j \) symbols after then of \( \sigma \), so we may write

\[ w(\sigma) = \sum_{j \geq 0} \frac{\mu(S_{\sigma,j})}{\mu([\sigma])} t^j \] (2)

where \( S_{\sigma,j} \) denotes the subset of \( [\sigma] \) containing an element \( \tau \in S \) that begins in \( \sigma \) and ends \( j \) symbols after the end of \( \sigma \). Observe that if \( \sigma \) ends in an element of \( S \), then we have \( S_{\sigma,0} = [\sigma] \) and hence \( w(\sigma) \geq 1 \). In general, the weight \( w(\sigma) \) is a measure of how close \( \sigma \) is to ending in an element of \( S \). The strategy here is study how \( w \) changes as you extend \( \sigma \) to the right by computing its weighted averages, and then use the results to bound from below the number of \( S \)-free extensions of \( \sigma \) and ultimately the entropy of \( \Sigma(S) \).

Define

\[ p_\sigma = \sum_{\tau \in S} \frac{\mu([\sigma \tau])}{\mu([\sigma])} t^{||\tau||} \]

Lemma \[ \] furnishes the upper bound

\[ p_\sigma \leq DA^{-1} \sum_{\tau \in S} (t/\lambda)^{||\tau||} = DA^{-1} p(t/\lambda) \] (3)

which is independent of \( \sigma \).

**Lemma 2.** If \( \sigma \) does not end in an element of \( S \), then

\[ \frac{1}{\mu([\sigma])} \sum_i \mu([\sigma i]) w(\sigma i) = \frac{w(\sigma) + p_\sigma}{t} \]

**Proof.** Using (2),

\[ \frac{1}{\mu([\sigma])} \sum_i \mu([\sigma i]) w(\sigma i) = \frac{1}{\mu([\sigma])} \sum_i \sum_j \mu(S_{\sigma i,j}) t^j \] (4)

An element of \( S_{\sigma i,j} \) has a unique \( \tau \in S \) ending \( j \) symbols after \( \sigma i \) and beginning within \( \sigma i \). This \( \tau \) can either begin within \( \sigma \) or begin at the final symbol \( i \), and accordingly we may decompose \( S_{\sigma i,j} = A_{\sigma i,j} \cup B_{\sigma i,j} \). Now

\[ \bigsqcup_i A_{\sigma i,j} = S_{\sigma,j+1} \]

and

\[ \bigsqcup_i B_{\sigma i,j} = \bigsqcup_{\tau \in S} [\sigma \tau] \]

are both clear from the definitions. Thus (4) is equal to

\[ \sum_j \left( \frac{\mu(S_{\sigma,j+1})}{\mu([\sigma])} + \sum_{\tau \in S} \frac{\mu([\sigma \tau])}{\mu([\sigma])} \right) t^j = \frac{w(\sigma) + p_\sigma}{t} \]

Note that the last equality relies on the fact that \( \sigma \) does not end in an element of \( S \), so the apparently missing \( \mu(S_{\sigma,0}) \) in the sum on the left vanishes.
3 Bounding entropy

Fix some $t > 1$ for the moment and let $\sigma$ be an $S$-free word with $w(\sigma) < 1$. We say that a word $\eta$ is good if $\sigma \eta$ is admissible and every intermediate word between $\sigma$ and $\sigma \eta$ (inclusive) has weight $w < 1$. In particular, $\sigma \eta$ is $S$-free if $\eta$ is good, since words ending in an element of $S$ have weight $\geq 1$. For a positive integer $m$, set

$$G(\sigma, m) = \bigsqcup_{\eta \text{ good} \atop |\eta| = m} [\sigma \eta]$$

Lemma 3. Suppose $\frac{1 + \rho t}{t} < r < 1$ for all words $\rho$, and let $\sigma$ be $S$-free with $w(\sigma) < 1$. We have

$$\frac{\mu(G(\sigma, m))}{\mu([\sigma])} \geq (1 - r)^m$$

for all $m \geq 1$.

Proof. Since extensions $\sigma i$ that end in an element of $S$ have $w(\sigma i) \geq 1$, we have

$$\sum_i \mu([\sigma i]) w(\sigma i) \geq \mu([\sigma]) - \mu(G(\sigma, 1))$$

Thus

$$\frac{\mu(G(\sigma, 1))}{\mu([\sigma])} \geq 1 - \frac{1}{\mu([\sigma])} \sum_i \mu([\sigma i]) w(\sigma i) = 1 - \left( \frac{w(\sigma) + \rho t}{t} \right) \geq 1 - r$$

which establishes the case $m = 1$.

Suppose the statement holds for some $m \geq 1$ and all $\sigma$. Observe that

$$G(\sigma, m + 1) = \bigsqcup_{\eta \text{ good} \atop |\eta| = m} G(\sigma \eta, 1)$$

For good $\eta$, the word $\sigma \eta$ is $S$-free and has $w(\sigma \eta) < 1$, so the base case and induction hypothesis give

$$\mu(G(\sigma, m + 1)) = \sum_{\eta \text{ good} \atop |\eta| = m} \mu(G(\sigma \eta, 1)) \geq \sum_{\eta \text{ good} \atop |\eta| = m} (1 - r) \mu([\eta]) = (1 - r) G(\sigma, m) \geq (1 - r)^{m+1} \mu([\sigma])$$

which establishes case $m + 1$. $\square$

Proposition 1. Suppose $r = \frac{1 + DA^{-1} \mu(t/\lambda)}{t} < 1$. We have $h(\Sigma) - h(\Sigma(S)) < - \log(1 - r)$

Proof. Using (3), we may apply Lemma 3 to the empty word $\sigma_0$ and conclude $\frac{\mu(G(\sigma_0, m))}{\mu([\sigma_0])} \geq (1 - r)^m$. The set $[\sigma_0]$ is simply $\Sigma$, but we retain $\sigma_0$ below for clarity. If $g$ denotes the number of good $\eta$ of length $m$, then we have

$$G(\sigma_0, m) = \sum_{\eta \text{ good} \atop |\eta| = m} \mu([\sigma_0]) \leq g DA^{-1} \lambda^{-m}$$

by Lemma 3. Thus we have produced for every $m \geq 1$ at least

$$g \geq AD^{-1} \lambda^m (1 - r)^m$$

words of length $m$ that are $S$-free, which implies that the entropy of $\Sigma(S)$ is at least

$$\lim_{m \to \infty} \frac{\log(AD^{-1} \lambda^m (1 - r)^m)}{m} = \log(\lambda) + \log(1 - r)$$

Since $\Sigma$ has entropy $\log(\lambda)$, this is the desired result. $\square$
4 Blocking

The condition \( r = \frac{1+DA^{-1}p(t/\lambda)}{t} < 1 \) in Proposition 1 implies \( t > 1 \) but also effectively limits \( t \) from above to roughly \( \lambda \). This in turn bounds \( r \) from below, limiting the direct utility of Proposition 1. The solution is to work with blocks of elements in \( \Sigma \). For each \( k \geq 1 \) let \( \Sigma_k \) denote the SFT on the alphabet of admissible words of length \( k \) in \( \Sigma \), where the transition \( |x_1 \cdots x_k|y_1 \cdots y_k| \) is admissible in \( \Sigma_k \) if and only if \( x_ky_1 \) is admissible in \( \Sigma \). Concatenating blocks furnishes a natural bijection \( \Sigma_k \to \Sigma \) that intertwines the shift map on \( \Sigma_k \) with the \( k \)th power of the shift map on \( \Sigma \). Accordingly, we have \( h(\Sigma_k) = kh(\Sigma) \).

To use the technique of the previous section, we must translate the collection \( S \) of forbidden words into an equivalent collection \( S_k \) of words in \( \Sigma_k \) - that is, one that cuts out the same subshift under the above bijection. In the process, we will also bound the associated polynomial

\[
p_k(t) = \sum_{\tau \in S_k} t^{\ell(\tau)}
\]

Let \( \tau \in S \) have length \( \ell \), suppose that \( k \leq \ell \), and write \( \ell = kq + r \) according to the division algorithm. To determine a collection of words in \( \Sigma_k \) that forbids \( \tau \) in \( \Sigma \), we must consider each of the \( k \) ways of tiling over \( \tau \) by blocks of length \( k \), according to the \( k \) possible positions of the beginning of \( \tau \) in the first block. Of these \( k \) positions, \( r + 1 \) require \( b = \lfloor \ell/k \rfloor \) blocks to tile over \( \tau \). Here, \( kb - \ell \) coordinates remain unspecified by \( \tau \), which means that we have at most \( B\lambda^{kb-\ell} \) words to consider at this position by (1). The remaining \( k - r - 1 \) positions require \( b + 1 \) blocks to tile over \( \tau \) and leave \( k(b+1) - \ell \) free coordinates.

\[
\begin{array}{c}
\tau \\
\begin{array}{c}
\begin{array}{c}
\text{\( b \) blocks}
\end{array}
\end{array}
\end{array}
\quad
\begin{array}{c}
\tau \\
\begin{array}{c}
\begin{array}{c}
\begin{array}{c}
\text{\( b + 1 \) blocks}
\end{array}
\end{array}
\end{array}
\end{array}
\quad
\begin{array}{c}
\square
\end{array}
\end{array}
\]

The total contribution to \( p_k(t) \) of the words associated to \( \tau \) is thus at most

\[
(r + 1)B\lambda^{kb-\ell}t^b + (k - r - 1)B\lambda^{k(b+1)-\ell}t^{b+1}
\]

Assuming that \( 1 \leq t \leq \lambda^k \), the contribution of \( \tau \) to \( p_k(t/\lambda^k) \) is then at most

\[
(r + 1)B\lambda^{-\ell}t^b + (k - r - 1)B\lambda^{-\ell}t^{b+1} \leq (r + 1)B\lambda^{-\ell}t^{b-1} + (k - r - 1)B\lambda^{2k-\ell}t^{b-1}
\]

Summing over \( \tau \in S \), we have

\[
p_k(t/\lambda^k) \leq kB\lambda^{2k} \sum_{\tau \in S} \left( \frac{t^{1/k}}{\lambda} \right)^{|\tau|} = kB\lambda^{2k}p \left( \frac{t^{1/k}}{\lambda} \right)
\]

(5)

**Proof of Theorem 1.** The constants \( A, B, \) and \( D \) depend on the underlying shift \( \Sigma \) but do not change upon replacing \( \Sigma \) by \( \Sigma_k \). Set \( C = DA^{-1}B \) and suppose that

\[
r = \frac{1 + kC\lambda^{2k}p(t^{1/k}/\lambda)}{t} < 1
\]

We may apply Lemma 3 as in the previous section but now to \( \Sigma_k \) to create at least \( AD^{-1}\lambda^{km}(1-r)^m \) words of length \( m \) in \( \Sigma_k \), and thus words of length \( km \) in \( \Sigma \), that avoid \( S \). The entropy of \( \Sigma(S) \) is therefore at least

\[
\lim_{m \to \infty} \frac{\log(AD^{-1}\lambda^{km}(1-r)^m)}{km} = \log(\lambda) + \frac{\log(1-r)}{k}
\]
as desired. \( \square \)
5 Growing words

Let $\ell$ denote the minimal length of an element of $S$. Then
\[ p(t^{1/k}/\lambda) \leq |S|(t^{1/k}/\lambda)^\ell \]
for $t \in (1, \lambda^k)$, so we consider
\[ r = \frac{1 + kC|S|\lambda^{2k-\ell}t^k}{t} \]
This function is minimized at
\[ t_{\text{min}} = \lambda^k \frac{\lambda^{-2k}/\ell}{((\ell - k)C|S|)^{k/\ell}} \]
and has minimum
\[ r_{\text{min}} = C^{k/\ell}|S|^{k/\ell} k/\ell \left(1 - \frac{k}{\ell}\right)^{k/\ell-1} \lambda^{-k+2k^2/\ell} \]
Note that $t_{\text{min}} \in (1, \lambda^k)$ as long as
\[ 1 < (\ell - k)C|S| < \lambda^{\ell-2k} \]
Let $\alpha \in (0, 1/2)$ and set $k = \lfloor \alpha \ell \rfloor$. Simple estimates show
\[ r_{\text{min}} = O(|S|^{\alpha} \lambda^{-\alpha(1-2\alpha)}) \]
Since $\lambda^{\ell-2k} \geq \lambda^{(1-2\alpha)}$, the condition (6) is satisfied as long as
\[ 1 < (\ell - k)C|S| < \lambda^{(1-2\alpha)} \]
Suppose that $|S|$ is bounded as $\ell \to \infty$. Then (7) holds for $\ell$ sufficiently large, and we have $r_{\text{min}} \to 0$. Since $-\log(1 - x) = O(x)$ for small $x$, Theorem 1 gives
\[ h(\Sigma) - h(\Sigma(S)) = O(\ell^{n-1} \lambda^{-\alpha(1-2\alpha)}) \]
Setting $\alpha = 1/4$ gives the best such bound, namely $O(\ell^{-3/4} \lambda^{-1/8})$, though we note that it is possible to improve this result slightly by using more refined estimates for $p_k$ in the previous section.
Now suppose that $|S|$ may be growing but subject to $|S| = O(\kappa^\ell)$ for some $\kappa < \lambda$. If we choose $\alpha$ small enough so that $\kappa < \lambda^{1-2\alpha}$, then condition (7) is satisfied for $\ell$ sufficiently large. We have
\[ r_{\text{min}} = O(\kappa^{\alpha} \ell^{\alpha} \lambda^{-\alpha(1-2\alpha)}) = O\left(\frac{\kappa}{\lambda^{1-2\alpha}} \ell^{\alpha}\right) \to 0 \]
as $\ell \to \infty$, so we may once again apply Theorem 1 to obtain
\[ h(\Sigma) - h(\Sigma(S)) = O\left(\frac{\kappa}{\lambda^{1-2\alpha}} \ell^{\alpha-1}\right) \to 0 \]
thereby establishing Theorem 2.

References
[1] D. A. Lind. Perturbations of shifts of finite type. SIAM J. Discrete Math., 2(3):350–365, 1989.
[2] Joseph S. Miller. Two notes on subshifts. Proc. Amer. Math. Soc., 140(5):1617–1622, 2012.
[3] Nick Ramsey. Perturbing subshifts of finite type: two words. unpublished, 2018.