Entropy bounds for multi-word perturbations of subshifts

NICK RAMSEY

*Department of Mathematical Sciences, DePaul University, 2320 N Kenmore Ave,
Chicago, IL 60614, USA
(e-mail: nramsey@depaul.edu)*

Abstract. Given a subshift Σ of finite type and a finite set S of finite words, let $\Sigma \langle S \rangle$ denote
the subshift of Σ that avoids S. We establish a general criterion under which we can bound
the entropy perturbation $h(\Sigma) - h(\Sigma \langle S \rangle)$ from above. As an application, we prove that
this entropy difference tends to zero with a sequence of such sets $S_1, S_2, \ldots$ under various
assumptions on the S_i.

Contents

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 Σ 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 $i \in \mathbb{Z}$. Recall
that the largest eigenvalue, λ, of T is related to the shift entropy of Σ by $\lambda = e^{h(\Sigma)}$, and
we assume throughout that $\lambda > 1$. Let S denote a finite set of finite admissible non-empty
words none of which contains another, and let $\Sigma \langle S \rangle$ denote the subshift of Σ consisting of
those elements of Σ in which none of the words in S appear.

We are interested in the entropy of the perturbed subshift $\Sigma\langle S\rangle$, and in particular seek to bound the entropy difference $h(\Sigma) - h(\Sigma\langle S\rangle)$. In [4], Lind addressed this problem in the case where S consists of a single word, and proved that the entropy of $\Sigma\langle S\rangle$ approaches that of Σ as the length ℓ of the word tends to infinity. He showed, moreover, that the difference in entropy is of exact order $O(\lambda^{-\ell})$ in the sense that it is bounded above and below by a constant multiple of $\lambda^{-\ell}$. Lind's entropy bounds stem from a lower bound on the *correlation polynomial* associated to the word in S. Such polynomials were studied by Guibas and Odlyzko in a series of papers [1–3], where they attribute the original definition to J. H. Conway.

In [8], the author adapted Lind's method to the case where S has more than one word. Here, one can define an $|S| \times |S|$ matrix of correlation polynomials analogous to the correlation polynomial for a single word, and an entropy bound comes down to a certain lower bound on the determinant of this matrix. 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 Σ in the case where S consists of two words of length tending to infinity. The entropy perturbation in this case is shown to be of order at most $O(\lambda^{-\ell})$ where ℓ is the length of the shorter word. Here again this must be the exact order, as can be seen by throwing out the longer word and appealing to Lind's lower bound for the remaining word of length ℓ. The correlation polynomial determinant seems to be a very natural quantity in this setting and deserving of further study with an eye towards adapting Lind's method more broadly, as it seems to give very sharp bounds. We note also that Pavlov in [7] has proven a result analogous to Lind's one-word result in the setting of multidimensional shifts.

In [5], Miller considered the following related problem: given a finite set S of finite non-empty words in a finite alphabet, determine whether there exists a 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. Miller defines

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

and establishes a simple numerical criterion involving $p(t)$ that ensures that his perturbation is non-empty. In this paper, we adapt Miller's method to a general subshift of finite type (SFT) and refine it to get lower bounds on the entropy of $\Sigma\langle S\rangle$, and thus upper bounds on the entropy perturbation. Our main result is as follows.

THEOREM 1. *There exists a constant $C > 0$ depending only on Σ 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*

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

then

$$h(\Sigma) - h(\Sigma\langle S\rangle) \leq -\frac{\log(1 - r)}{k}.$$

This theorem is shown to have consequences for the 'growing words' problems above. A perturbation bound of $O(\lambda^{-\ell})$ seems beyond the method, but we can establish

$O(\ell^{-3/4}\lambda^{-\ell/8})$ when S consists of any fixed number of words of minimal length $\ell \to \infty$. The method can also be applied to a growing set of words of increasing length with sufficient control over the growth, as illustrated by the following result. Suppose that $S_1, S_2, \ldots$ is a sequence of sets as above, let ℓ_i denote the minimal length in S_i, and suppose $\ell_i \to \infty$.

THEOREM 2. *Suppose there exists $\kappa < \lambda$ such that $|S_i| = O(\kappa^{\ell_i})$ as $i \to \infty$. Then*

$$h(\Sigma) = \lim_{i\to\infty} h(\Sigma \langle S_i \rangle).$$

2. *Parry measure and the weight function $w(\sigma)$*
Let u and v denote left and right λ-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 Σ, 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. In [6], Parry defined a shift-invariant measure of maximal entropy μ on Σ that can be characterized on non-empty cylinder sets by

$$\mu([i x_1 x_2 \cdots x_{k-1} j]) = \frac{u_i v_j}{\lambda^k}.$$

Note that the notation $[\sigma]$ only defines a cylinder set up to shifts. Where it is important to have an actual set to work with (e.g., in defining f_σ below) we take σ to begin at coordinate 0 in forming $[\sigma]$. If σ fails to be admissible, then $[\sigma]$ is taken to be empty.

Let $N(\sigma, k)$ denote the number of words η of length k such that $\sigma \eta$ is admissible. Since T is irreducible, it follows from Perron–Frobenius theory that there exist positive constants A, B such that

$$A\lambda^k \leq N(\sigma, k) \leq B\lambda^k \tag{1}$$

for all words σ 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 σ, τ.

Proof. We may assume that $\sigma\tau$ is admissible, for otherwise the result is trivial. The explicit description of μ on cylinder sets implies that D is an upper bound for the ratio among the $\mu([\sigma\eta])$ as η varies among the words of a given positive length for which $\sigma\eta$ is admissible. Thus,
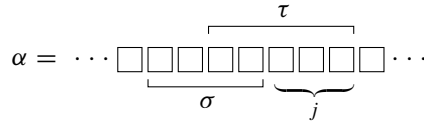
$$\lambda^{|\tau|}\mu([\sigma\tau]) \leq A^{-1}N(\sigma, |\tau|)\mu([\sigma\tau]) = A^{-1}\sum_\eta \mu([\sigma\tau])$$

$$\leq A^{-1}\sum_\eta D\mu([\sigma\eta])$$

$$= DA^{-1}\mu([\sigma])$$

where the sums are taken over those η with $|\eta| = |\tau|$ and $\sigma\eta$ admissible. $\square$

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

$$f_\sigma(\alpha) = \sum_{\tau \in S} \sum_j t^j$$

where the inner sum is over $j \geq 0$ such that τ occurs in α beginning within σ and ending j symbols beyond the end of σ:

Observe that the function f_σ is locally constant on $[\sigma]$. We define a weight function on admissible words by

$$w(\sigma) = \frac{1}{\mu([\sigma])} \int_{[\sigma]} f_\sigma \, d\mu.$$

Note that the empty word σ_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 σ, so we may write

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

where $S_{\sigma,j}$ denotes the subset of $[\sigma]$ containing an element $\tau \in S$ that begins in σ and ends j symbols after the end of σ. Observe that if σ 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 σ is to ending in an element of S. The strategy here is study how w changes as you extend σ to the right by computing its weighted averages, and then use the results to bound from below the number of S-free extensions of σ and ultimately the entropy of $\Sigma\langle S \rangle$.

Define

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

Lemma 1 furnishes the upper bound

$$p_\sigma \leq D A^{-1} \sum_{\tau \in S} (t/\lambda)^{|\tau|} = D A^{-1} p(t/\lambda) \tag{3}$$

which is independent of σ.

LEMMA 2. *If σ 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),

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

An element of $S_{\sigma i, j}$ has a unique $\tau \in S$ ending j symbols after σi and beginning within σi. This τ can begin either within σ or at the final symbol i, and accordingly we may decompose $S_{\sigma i, j} = A_{\sigma i, j} \sqcup B_{\sigma i, j}$. Explicitly, the set $A_{\sigma i, j}$ consists of elements of $[\sigma i]$ that contains a word in S beginning within σ and ending j symbols after σi. The set $B_{\sigma i, j}$ consists of elements of $[\sigma i]$ in which σ is immediately followed by an element of S of length $j + 1$. Observe that

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

and

$$\bigsqcup_i B_{\sigma i, j} = \bigsqcup_{\substack{\tau \in S \\ |\tau| = j+1}} [\sigma \tau]$$

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

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

Note that the last equality relies on the fact that σ 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 σ be an S-free word with $w(\sigma) < 1$. We say that a word η is *good* if $\sigma \eta$ is admissible and every intermediate word between σ and $\sigma \eta$ (inclusive) has weight $w < 1$. In particular, $\sigma \eta$ is S-free if η is good, since words ending in an element of S have weight greater than or equal to 1. For a positive integer m, set

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

LEMMA 3. *Suppose $(1 + p_\rho)/t < r < 1$ for all words ρ, and let σ 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 σi 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,

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

which establishes the case $m = 1$.

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

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

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

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

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

which establishes case $m+1$. □

PROPOSITION 1. *Suppose $r = (1 + DA^{-1}p(t/\lambda))/t < 1$. We have $h(\Sigma) - h(\Sigma\langle S\rangle) \leq -\log(1-r)$.*

Proof. Using (3), we may apply Lemma 3 to the empty word σ_0 and conclude $\mu(G(\sigma_0, m))/\mu([\sigma_0]) \geq (1-r)^m$. The set $[\sigma_0]$ is simply Σ, but we retain σ_0 below for clarity. If g denotes the number of good η of length m, then by Lemma 1 we have

$$\frac{G(\sigma_0, m)}{\mu([\sigma_0])} = \sum_{\eta} \frac{\mu([\sigma_0\eta])}{\mu([\sigma_0])} \leq gDA^{-1}\lambda^{-m}$$

where the sum is over such η. 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 $\Sigma\langle S\rangle$ is at least

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

Since Σ has entropy $\log(\lambda)$, this is the desired result. □

4. *Blocking*

The condition $r = (1 + DA^{-1}p(t/\lambda))/t < 1$ in Proposition 1 implies $t > 1$ but also effectively limits t from above to roughly λ. This in turn bounds r from below, limiting the direct utility of Proposition 1. The solution is to work with blocks of elements in Σ. For each $k \geq 1$ let Σ_k denote the SFT on the alphabet of admissible words of length k in Σ, where the transition $[x_1 \cdots x_k][y_1 \cdots y_k]$ is admissible in Σ_k if and only if $x_k y_1$ is admissible in Σ. Concatenating blocks furnishes a natural bijection $\Sigma_k \longrightarrow \Sigma$ that intertwines the shift map on Σ_k with the kth power of the shift map on Σ. 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 Σ_k, that is, one that cuts out the same subshift under the above bijection. In the process, we will also bound the associated polynomial

$$p_k(t) = \sum_{\zeta \in S_k} t^{|\zeta|}.$$

Let $\tau \in S$ have length ℓ, suppose that $k \leq \ell$, and write $\ell = kq + s$ with $0 \leq s < k$ according to the division algorithm. To determine a collection of words in Σ_k that forbids τ in Σ, we must consider each of the k ways of tiling over τ by blocks of length k, according to the k possible positions of the beginning of τ in the first block. Of these k positions, $s + 1$ require $b = \lceil \ell/k \rceil$ blocks to tile over τ. Here, $kb - \ell$ coordinates remain unspecified by τ, which means that we have at most $B\lambda^{kb-\ell}$ words to consider at this position by (1). The remaining $k - s - 1$ positions require $b + 1$ blocks to tile over τ and leave $k(b+1) - \ell$ free coordinates.
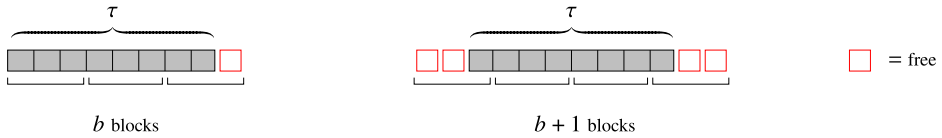
The total contribution to $p_k(t)$ of the words associated to τ is thus at most

$$(s+1)B\lambda^{kb-\ell}t^b + (k-s-1)B\lambda^{k(b+1)-\ell}t^{b+1}.$$

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

$$(s+1)B\lambda^{-\ell}t^b + (k-s-1)B\lambda^{-\ell}t^{b+1} \leq (s+1)B\lambda^{k-\ell}t^{b-1} + (k-s-1)B\lambda^{2k-\ell}t^{b-1}$$
$$\leq kB\lambda^{2k-\ell}t^{\ell/k}.$$

Summing over $\tau \in S$, we have

$$p_k(t/\lambda^k) \leq kB\lambda^{2k} \sum_{\tau \in S} \left(\frac{t^{1/k}}{\lambda}\right)^{|\tau|} = kB\lambda^{2k} p\left(\frac{t^{1/k}}{\lambda}\right). \tag{5}$$

Proof of Theorem 1. The constants A, B, and D depend on the underlying shift Σ but do not change upon replacing Σ by Σ_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 Σ_k to create at least $AD^{-1}\lambda^{km}(1-r)^m$ words of length m in Σ_k, and thus words of length km in Σ, that avoid S. The entropy of $\Sigma\langle S \rangle$ 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 ℓ 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

$$r = \frac{1 + kC|S|\lambda^{2k-\ell} t^{\ell/k}}{t}.$$

This function is minimized at

$$t_{\min} = \lambda^k \frac{\lambda^{-2k^2/\ell}}{((\ell - k)C|S|)^{k/\ell}}$$

and has minimum

$$r_{\min} = C^{k/\ell}|S|^{k/\ell}\ell^{k/\ell}\left(1 - \frac{k}{\ell}\right)^{k/\ell-1}\lambda^{-k+2k^2/\ell}.$$

Note that $t_{\min} \in (1, \lambda^k)$ as long as

$$1 < (\ell - k)C|S| < \lambda^{\ell-2k}. \tag{6}$$

Let $\alpha \in (0, 1/2)$ and set $k = \lfloor \alpha\ell \rfloor$. Simple estimates show

$$r_{\min} = O(|S|^\alpha \ell^\alpha \lambda^{-\ell\alpha(1-2\alpha)}).$$

Since $\lambda^{\ell-2k} \geq \lambda^{\ell(1-2\alpha)}$, condition (6) is satisfied as long as

$$1 < (\ell - k)C|S| < \lambda^{\ell(1-2\alpha)}. \tag{7}$$

Suppose that $|S|$ is bounded as $\ell \to \infty$. Then (7) holds for ℓ sufficiently large, and we have $r_{\min} \to 0$. Since $-\log(1 - x) = O(x)$ for small x, Theorem 1 gives

$$h(\Sigma) - h(\Sigma\langle S \rangle) = O(\ell^{\alpha-1}\lambda^{-\ell\alpha(1-2\alpha)}).$$

Setting $\alpha = 1/4$ gives the best such bound, namely $O(\ell^{-3/4}\lambda^{-\ell/8})$, 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 α small enough so that $\kappa < \lambda^{1-2\alpha}$, then condition (7) is satisfied for ℓ sufficiently large. We have

$$r_{\min} = O(\kappa^{\alpha\ell}\ell^\alpha\lambda^{-\ell\alpha(1-2\alpha)}) = O\left(\left(\frac{\kappa}{\lambda^{1-2\alpha}}\right)^{\alpha\ell}\ell^\alpha\right) \to 0$$

as $\ell \to \infty$, so we may once again apply Theorem 1 to obtain

$$h(\Sigma) - h(\Sigma\langle S \rangle) = O\left(\left(\frac{\kappa}{\lambda^{1-2\alpha}}\right)^{\alpha\ell}\ell^{\alpha-1}\right) \to 0,$$

thereby establishing Theorem 2.

References

[1] L. J. Guibas and A. M. Odlyzko. Maximal prefix-synchronized codes. *SIAM J. Appl. Math.* **35**(2) (1978), 401–418.
[2] L. J. Guibas and A. M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *J. Combin. Theory Ser. A* **30**(2) (1981), 183–208.

[3] L. J. Guibas and A. M. Odlyzko. Periods in strings. *J. Combin. Theory Ser. A* **30**(1) (1981), 19–42.
[4] D. A. Lind. Perturbations of shifts of finite type. *SIAM J. Discrete Math.* **2**(3) (1989), 350–365.
[5] J. S. Miller. Two notes on subshifts. *Proc. Amer. Math. Soc.* **140**(5) (2012), 1617–1622.
[6] W. Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.* **112** (1964), 55–66.
[7] R. Pavlov. Perturbations of multidimensional shifts of finite type. *Ergod. Th. & Dynam. Sys.* **31**(2) (2011), 483–526.
[8] N. Ramsey. Perturbing subshifts of finite type: two words. *Preprint*, 2019, arXiv:1902.03352.