SCHRÖDINGER OPERATORS
WITH POTENTIALS GENERATED BY PRIMITIVE
SUBSTITUTIONS: AN INVITATION

BY DAVID DAMANIK

Abstract. This paper demonstrates in a step by step manner how the spectral
properties of Schrödinger operators with potentials generated by primitive sub-
stitutions derive from the underlying algebraic structure and tries to introduce
the non-specialist to this interesting field.

1. Introduction. The interest in Schrödinger operators with potentials
generated by primitive substitutions which has been growing since the late
80's is coming from at least two different sources. Firstly, quasi-crystals phys-
ically motivated this study because they exhibit self-similarities which can be
described by means of substitution rules. Secondly, there has been a large
interest in such operators from a mathematical point of view because of their
interesting spectral properties. It seems to be the rule that they have purely
singular continuous spectrum supported on a Cantor set of Lebesgue measure
zero. This belief is supported by a general approach using a certain dynamical
system, the trace map, which is more or less directly induced by the substitu-
tion rule.

The mathematical interest is mainly coming from one heavily studied ex-
ample, the Fibonacci hamiltonian. The papers [1,2,3] still consider this op-
erator in its original quasiperiodic setting and prove most of its properties
mentioned above. The situation in this quasiperiodic setting was improved
and generalized by the authors of [5] (see also [4]) who first used the strat-
ey to be sketched below. Another possible generalization, namely extension
of the strategy from [5] to the class of Schrödinger operators with potentials
generated by substitutions, was worked out in [11] after it was realized that it applies to other examples from this class, too [7,9].

The aim of this paper is to introduce this general approach to the interested reader who has no familiarity with this class of operators so far. Therefore, we have tried to motivate and to define the necessary quantities in great detail and to show step by step how this canonical method works. The organization is as follows: section 2 introduces substitutions on finite alphabets and substitution sequences which are simply fixpoints of homomorphically extended substitutions. These substitution sequences give potentials for discrete one-dimensional Schrödinger operators as described in section 3. Some interesting properties of these operators are discussed in section 4 together with possible applications. In section 5 it is shown how all the spectral properties except the absence of eigenvalues follow by proving one key equality. The point spectrum is then discussed in section 6. Finally, section 7 states some open problems.

There is another introductory paper [14] by Bovier and Ghez emphasizing on reporting the recent results concerning substitution hamiltonians in order to rectify some claims which strictly contradict results already proven rigorously.

2. Substitutions and induced dynamical systems. This section introduces the concepts of substitutions and substitution sequences and describes how the underlying ergodic flow of the Schrödinger operator to be defined in the next section arises from a substitution sequence.

DEFINITION 2.1. An alphabet is a finite set \( A = \{a_1, \ldots, a_r\} \) of symbols \( a_i \). The formal product \( A^k \) contains the words of length \( k \) over \( A \). \( A^* \equiv \cup_{k \in \mathbb{N}} A^k \) is the set of all words over \( A \). The length of a word \( w = b_1 \ldots b_s \in A^* \) is given by \( |w| \equiv s \).

DEFINITION 2.2. A substitution is a mapping \( \zeta : A \rightarrow A^* \) where \( A \) is an alphabet.

As the name indicates, starting with a word over \( A \) the mapping \( \zeta \) will be used to obtain a new word simply by substituting symbol by symbol. \( \zeta \) is extended homomorphically to mappings \( \zeta : A^* \rightarrow A^* \) and \( \zeta : A^N \rightarrow A^N \) which will still be denoted by \( \zeta \).

DEFINITION 2.3. \( s \in A^N \) is called substitution sequence if it is a fixpoint of \( \zeta \).

Because substitution sequences induce dynamical systems which in turn give rise to Schrödinger operators describing the dynamics for structures with self-similarity properties governed by the substitution, it is an interesting question how the existence of substitution sequences can be ensured. In the fol-
lowing, $A^N$ and $A^Z$ will always be equipped with the product topology, i.e.
the topology of pointwise convergence. One checks:

**Proposition 2.4.** Suppose the two following conditions are satisfied:

(i) There exists a symbol $a \in A$ such that $\zeta(a) = aw$ with a suitably chosen
    $w \in A^*$, i.e. $\zeta(a)$ begins with $a$.

(ii) $|\zeta^n(a)| \to \infty$ for $n \to \infty$ and the symbol $a$ from condition (i).

Then the sequence $(\zeta^n(a))$ converges in $A^N$ and the limit $\zeta^\infty(a)$ is a fixpoint
of $\zeta$.

We are now about to consider an example. Actually, it will serve as a
running example for the paper because the substitution to be introduced be-
low yields the most heavily studied Schrödinger operator with a substitution
potential.

**Example 2.5.** Consider $A = \{a, b\}$, $\zeta_F(a) = ab$, $\zeta_F(b) = a$. The iterates are

\[
\begin{align*}
\zeta^0_F(a) &= a \\
\zeta^1_F(a) &= ab \\
\zeta^2_F(a) &= aba \\
\zeta^3_F(a) &= abaab \\
\zeta^4_F(a) &= abaababa \\
\zeta^5_F(a) &= abaababaabaab \\
& \\
\vdots
\end{align*}
\]

Obviously, $\zeta^n_F(a) = \zeta^{n-1}_F(a)\zeta^{n-2}_F(a)$ which follows easily from the substitution
rule. Hence, $s_F \equiv \zeta^\infty_F(a) = abaababaabaab\ldots$ is the substitution sequence
obtained from the procedure indicated in Proposition 2.4.

Now, a substitution sequence $s$ induces a dynamical system as follows: let $\bar{s}$
be an arbitrary element of $A^Z$ which coincides with $s$ on the positive integers.
Let $T$ denote the standard shift on $A^Z$: $(Tw)(n) \equiv w(n + 1)$. Finally, define
$\Omega$ to be the set of accumulation points of the sequence $(T^n(\bar{s}))$ in $A^Z$, i.e.

\[
\Omega = \left\{ \omega \in A^Z : \omega = \lim_{i \to \infty} T^{n_i}(\bar{s}) \text{ and } n_i \to \infty \right\}
\]

The extension from one-sided to two-sided sequences is necessary since we are going to consider Schrödinger operators in $l^2(Z)$ rather than in $l^2(N)$.

**Definition 2.6.** A substitution $\zeta$ is called *primitive* if there exists $k \in N$
such that for every pair of symbols $(a_i, a_j)$ the word $\zeta^k(a_i)$ contains the sym-
bol $a_j$. 
Obviously, the substitution $\zeta_F$ from Example 2.5 is primitive, choose $k = 2$. The substitution dynamical system $(\Omega, T)$ arising from a primitive substitution has the following remarkable properties [17]:

**Proposition 2.7.** Let $\zeta$ be primitive. Then:

(i) $(\Omega, T)$ is uniquely ergodic.
(ii) $(\Omega, T)$ is minimal.
(iii) Every word from $A^*$ that occurs in some $w \in \Omega$ occurs in every $v \in \Omega$ and even with a well defined frequency.

Unique ergodicity means that there is exactly one probability measure $P$ on $\Omega$ which makes the system $(\Omega, T, P)$ ergodic. Minimality holds if every orbit is dense, i.e. $\{T^n(w)\}_{n \in \mathbb{Z}} = \Omega$ for every $w \in \Omega$.

3. **Associated Schrödinger operators.** This section introduces Schrödinger operators with potentials generated by substitutions, shows how they fit into the class of ergodic operators, introduces the basic concept of transfer matrices and defines the trace map - a very useful tool in the investigation of the spectral properties of such operators.

Let $\zeta : A \to A^*$ be a primitive substitution and $\omega \in \Omega$ where $\Omega$ is induced by $\zeta$ as described in section 2. For a given $f : A \to \mathbb{R}$ let

$$V_\omega : \mathbb{Z} \to \mathbb{R}, V_\omega(n) \equiv f(\omega(n)).$$

Thus, symbols are simply replaced by real numbers. The sequence $V_\omega$ now serves as a potential for a discrete Schrödinger operator $H_\omega$ in $l^2(\mathbb{Z})$, namely

$$(H_\omega u)(n) \equiv u(n + 1) + u(n - 1) + V_\omega(n)u(n).$$

$\{H_\omega\}_{\omega \in \Omega}$ is an ergodic family [18] since the underlying flow $T$ on $\Omega$ is ergodic with respect to $P$, compare Proposition 2.7. Hence, there exist sets $\Sigma, \Sigma_{pp}, \Sigma_{sc}, \Sigma_{ac} \subseteq \mathbb{R}$ such that:

$$\sigma(H_\omega) = \Sigma, \sigma_{pp}(H_\omega) = \Sigma_{pp}, \sigma_{sc}(H_\omega) = \Sigma_{sc}, \sigma_{ac}(H_\omega) = \Sigma_{ac} \quad P - a.s.$$ 

Consider the eigenvalue equation for $H_\omega$:

$$(1) \quad u(n + 1) + u(n - 1) + V_\omega(n)u(n) = Eu(n).$$

As always in the investigation of one-dimensional Schrödinger operators, this equation is about to be reformulated in terms of unimodular $2 \times 2$ matrices - the transfer matrices $M_{\omega,E}(\cdot)$. Let $u$ be a solution of (1). Define

$$u(n) \equiv \begin{pmatrix} u(n+1) \\ u(n) \end{pmatrix}$$
and

\[ M_{\omega,E}(n) = \begin{cases} 
T_{\omega,E}(n) \cdots T_{\omega,E}(1) & n > 0 \\
Id & n = 0 \\
T_{\omega,E}(n+1)^{-1} \cdots T_{\omega,E}(0)^{-1} & n < 0 
\end{cases} \]

where

\[ T_{\omega,E}(n) = \begin{pmatrix} E - V_\omega(n) & -1 \\ 1 & 0 \end{pmatrix}. \]

It is easily checked that

\[ u(n) = M_{\omega,E}(n)u(0) \forall n \in \mathbb{Z}. \]

Thus, a solution of (1) is already determined by two consecutive values and the other values can be computed recursively. Moreover, the self-similarity of the potential which is of course due to the generating-by-substituting-process yields recursive relations for the transfer matrices:

**Example 3.1.** Let \( s_F \) be the Fibonacci sequence. By symmetric extension one obtains a two-sided sequence \( \tilde{s}_F \) which belongs to \( \Omega \) as we will explain below. The operator

\[ (H_F u)(n) \equiv u(n+1) + u(n-1) + V(n)u(n) \]

in \( l^2(\mathbb{Z}) \) where the potential \( V \) is coming from \( \tilde{s}_F \) via the function \( f(a) = 0, f(b) = 1 \) is called Fibonacci operator. The relation

\[ \zeta_F^n(a) = \zeta_F^{n-1}(a) \zeta_F^{n-2}(a) \]

yields the following recursion for the transfer matrices in a straightforward manner:

\[ M_E(F_n) = M_E(F_{n-2})M_E(F_{n-1}) \]

where \( F_n \) is the \( n \)-th Fibonacci number. Setting \( x_n(E) \equiv \frac{1}{2} tr(M_E(F_n)) \) and using \( det M_E(\cdot) = 1 \), we have

\[ x_n(E) = 2x_{n-1}(E)x_{n-2}(E) - x_{n-3}(E). \]

This equation can be regarded as a dynamical system on \( \mathbb{R}^3 \), namely iteration of the following function:

\[ TM : \mathbb{R}^3 \to \mathbb{R}^3, \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} y \\ z \\ 2zy - x \end{pmatrix} \]
(2) (3) resp.) is called the trace map. Its importance will become clear in section 5: the set

\[ B_\infty \equiv \{ E \in \mathbb{R} : (x_n(E))_{n \in \mathbb{N}} \text{ remains bounded} \} \]

which corresponds to the stable set of (2) actually coincides with the spectrum of \( H_F \)!

This strategy has been generalized to a certain class of substitutions in [11]. Intuitively, it is clear that the self-similarity of the substitution sequence results in recursive relations for the transfer matrices and therefore allows the definition of a trace map. Nevertheless, more complicated substitutions yield recursive relations far from being simple. Hence, the stable set might be much more difficult to determine and thus this approach is less straightforward in the general setting. However, the same strategy of proof is used, so that a good understanding of the Fibonacci case reveals all the ideas which lead to the general picture for such operators.

**Remark 3.2.** Basically, starting with a one-sided sequence there are two common methods how to produce a two-sided sequence (compare [13]):

(i) symmetric extension to the left
(ii) shifting to the left, accumulation points

The first method perfectly allows to use the trace map in one direction only since it is the same for both directions. This makes it easier to exclude eigenvalues. On the other hand, one is tempted to use the second method if there is a need for the ergodic framework (with its powerful implications, e.g. Kotani's theorem, compare Section 4). The (one-sided) ergodicity results from [17] then carry over to the two-sided case. The second method, however, has the disadvantage that in general it is not clear how to use trace map recursions for the transfer matrices because the origin where the recursion is centered is 'shifted away'. Thus, it would be very comfortable to combine the two methods if possible. Therefore, it is sufficient to look for both left and right fixpoints of \( \zeta \), that is \( \zeta^\infty(a_l), \zeta^\infty(a_r) \) where \( \zeta(a_l) \) ends with \( a_l \) and \( \zeta(a_r) \) begins with \( a_r \), and to look for an occurrence of \( a_la_r \) in \( \zeta^\infty(a_r) \) for instance. In this case, \( \zeta^\infty(a_l)\zeta^\infty(a_r) \in A^\mathbb{Z} \) would be an accessible sequence concerning trace map investigation which belongs to \( \Omega \). Obviously, this procedure applies to the Fibonacci substitution \( \zeta_F \) (with \( a_l = a_r = a \) and \( \zeta_F^2 \) instead of \( \zeta_F \)).

**4. Constancy of the Lyapunov exponent.** This section consists of an extended remark about the Lyapunov exponent \( \gamma \) which is defined as follows. Let

\[ \gamma^\pm(E) \equiv \lim_{n \to \pm\infty} \frac{1}{|n|} \log \| M_{\omega,E}(n) \|. \]
For general ergodic operators, $\gamma^\pm(E)$ exist for fixed $E$ for $P$-almost all $\omega$, are independent of $\omega$ and satisfy

$$\gamma^+(E) = \gamma^-(E) = \gamma(E),$$

compare [18]. In addition, the Lyapunov exponent for Schrödinger operators with substitution potentials exists uniformly and is even constant over $\Omega$ [12]:

**Theorem 4.1.** Let $\Omega \subseteq \mathbb{A}^\mathbb{Z}$ be defined by a primitive substitution on a finite set $A$. Then for every $E \in \mathbb{C}$ there exists a $\gamma(E) \in \mathbb{R}$ such that

$$\gamma(E) = \lim_{n \to \infty} \frac{1}{n} \log \|M_{\omega,E}(n)\|$$

holds for all $\omega \in \Omega$.

The proof uses some results about subadditive set functions from [8]. Besides, the main ingredient is the observation from Proposition 2.7: two sequences from $\Omega$ are very similar, in the sense that the same words occur in the sequences and that they coincide on a given bounded set if they are sufficiently close in the topology of $\Omega$. Hence, the defining sequences of transfer matrices also coincide on this set.

This result is extremely useful because in many cases there is a concrete element of $\Omega$ which is more accessible for the computation of the Lyapunov exponent than other elements because the recursive relations for the transfer matrices hold. It is this concrete potential one is mainly interested in. The imbedding into the ergodic framework is done for the following reasons:

(i) the spectrum remains unchanged anyway

(ii) the ergodic framework is suggested by Proposition 2.7

(iii) the existence of the Lyapunov exponent for general ergodic operators

(iii) the constancy of the Lyapunov exponent

Thus, a very nice interplay between deterministic results and ‘random point of view’ is demonstrated. Moreover, this strategy also works the other way round: by investigating a concrete potential from the hull $\Omega$ one gets results for all potentials. In particular, we have the following important quantities as invariants:

(i) Lyapunov exponent

(ii) Lebesgue measure of the spectrum

(iii) gap structure of the spectrum

Now, there is another important invariant: the absolutely continuous spectrum. Actually, it is empty [15]:
Theorem 4.2. Let $\Omega \subseteq A^\mathbb{Z}$ be defined by a primitive substitution on a finite set $A$. Then: $\sigma_{ac}(H_\omega) = \emptyset$ for all $\omega \in \Omega$.

This theorem follows from Kotani theory [18] together with the observation in [6] which essentially states that an aperiodic ergodic Schrödinger operator with a potential which takes only finitely many values has $\Sigma_{ac} = \emptyset$ (with $\Sigma_{ac}$ as in section 3). Of course, in general $\sigma_{ac}(H_\omega) = \Sigma_{ac} P - \text{a.s.}$ only, but Theorem 4.2 follows if uniform existence and constancy of the Lyapunov exponent is established, compare [15].

5. Lebesgue measure, Cantor property and singularity of the spectrum. The class of Schrödinger operators with potentials generated by primitive substitutions reveals the tendency of having Cantor spectrum of zero Lebesgue measure. These results are obtained by following a canonical strategy which will be presented and discussed in this section.

The ultimate goal is to establish the following equalities:

$$\sigma(H_\omega) = B_{\infty} = \{E \in \mathbb{R} : \gamma(E) = 0\}$$

Firstly, we would like to comment about the quantities:

(i) $B_{\infty}$ is the stable set for the trace map, compare section 3. The trace map is a deterministic quantity associated to a concrete element $\omega_s \in \Omega$ which represents the defining substitution sequence on the positive semi-axis.

(ii) $\sigma(H_\omega)$ is the same for all $\omega$, so that is is enough to study $\sigma(H_{\omega_s})$.

(iii) The Lyapunov exponent $\gamma$ is also the same for all $\omega$. Thus, it can be considered as the concrete Lyaponov exponent for $H_{\omega_s}$ as well as the Lyapunov exponent yielded by the ergodic approach.

Secondly, we would like to demonstrate how the following metatheorem follows from the above equalities:

Theorem 5.1. Suppose (4) holds. Then:

(i) $\sigma(H_\omega)$ is a Cantor set for all $\omega \in \Omega$

(ii) $\sigma(H_\omega)$ has zero Lebesgue measure for all $\omega \in \Omega$

(iii) $\sigma_{ac}(H_\omega) = \emptyset$ for all $\omega \in \Omega$

Since the potentials take only finitely many values, Kotani theory gives

$$|\{E \in \mathbb{R} : \gamma(E) = 0\}| = 0$$

unless the potential is periodic [6]. Hence, (ii) follows immediately. Now, absolutely continuous spectral measures cannot be carried on a set of zero
Lebesgue measure, so that \((iii)\) follows from \((ii)\). Next, since the operators under consideration cannot have isolated points in their spectrum \([18]\), all we have to show for \((i)\) is denseness of the gaps. But this also follows from \((ii)\).

Thus, all the results follow from \(\sigma(H_\omega) = \{E \in \mathbb{R} : \gamma(E) = 0\}\). The reason why \(B_\infty\) is linked between them in \((4)\) is simply that the trace map plays the role of a link between spectrum and Lyapunov exponent. So, finally we would like to indicate how \((4)\) can be proved:

(i) \(\sigma(H_\omega) \subseteq B_\infty\): This is shown by strong approximation. A potential coming from \(\omega\), yields periodic approximants simply by a cut-and-repeat procedure. Strong convergence is checked and the transfer matrices coincide on the interval of periodicity of the approximant.

(ii) \(B_\infty \subseteq \{E \in \mathbb{R} : \gamma(E) = 0\}\): The intuition that bounded traces should imply subexponential growth of the norms in the given situation is to be justified. Of course, this is the hardest part of the proof and it has caused Bovier and Ghez some troubles to achieve it in their general setting.

(iii) \(\{E \in \mathbb{R} : \gamma(E) = 0\} \subseteq \sigma(H_\omega)\): This follows from general results \([18]\), hence there is nothing to prove in our concrete situation.

6. The point spectrum. This section discusses the point spectrum of Schrödinger operators with substitution potentials. Unfortunately, so far there is no canonical approach yielding a general result as in the last section. However, it is believed that as a rule eigenvalues should be absent. The point spectrum has been investigated mainly for some prominent examples. Other results have been obtained by generalizing an idea from the proof of absence of eigenvalues for the Fibonacci operator and by trying to recover symmetries of the potential in the solutions of the eigenvalue equation. In all cases not only the absence of \(l^2\)-solutions has been shown – the solutions do not even tend to zero because of the recovered symmetries. In the following, we indicate how translation \([3,11]\) or reflection \([15]\) symmetry of the potential can be used to prove absence of eigenvalues.

6.1 Translation symmetry. Consider again the Fibonacci case. We would like to show:

\[
    u(2F_n) = M_E(F_n)^2 u(0).
\]

Then, the absence of eigenvalues follows from the following elementary observation \([3]\) because the traces \(x_n(E)\) are bounded on \(\sigma(H_F)\), compare \((4)\).

**Lemma 6.1.** Let \(B\) be a \(2 \times 2\) matrix with \(\det B = 1\). Then for all \(x \in \mathbb{C}^2:\)

\[
    \max \{|\text{tr } B| \cdot \|Bx\|, \|B^2x\|\} \geq \frac{1}{2}\|x\|.
\]
We know already

(i) \( s_F = \zeta_F(s_F) \)
(ii) \( \zeta_F^n(a) = \zeta_F^{n-1}(a)\zeta_F^{n-2}(a) \)
(iii) \( \zeta_F^n(a) \) is a prefix of \( s_F \).
(iv) \( |\zeta_F^n(a)| = F_{n+1} \)

where \( s_F \) is the substitution sequence obtained from \( \zeta_F \) by iteration on \( a \). So, for proving (5) we have to show that \( \zeta_F^{n-1}(a)\zeta_F^{n-1}(a) \) is a prefix of \( s_F \):

\[
\zeta_F^{n+1}(a) = \zeta_F^n(a)\zeta_F^{n-1}(a) \\
= \zeta_F^{n-1}(a)\zeta_F^{n-2}(a)\zeta_F^{n-2}(a)\zeta_F^{n-3}(a) \\
= \zeta_F^{n-1}(a)\zeta_F^{n-2}(a)\zeta_F^{n-3}(a)\zeta_F^{n-4}(a)\zeta_F^{n-4}(a)\zeta_F^{n-3}(a) \\
= \zeta_F^{n-1}(a)\zeta_F^{n-1}(a)\zeta_F^{n-4}(a)\zeta_F^{n-3}(a)
\]

This idea has been generalized in [11].

6.2 Reflection symmetry. A word is called a palindrome if it is the same when read backwards. A sequence \( \omega \in A^\mathbb{Z} \) is called strongly palindromic if there exists for every \( B > 0 \) a sequence \( w_i \) of palindromes of length \( l_i \) in \( \omega \) centered at \( m_i \to \infty \) such that \( e^{Bm_i}/l_i \to 0 \). Now, in [15] Hof, Knill and Simon prove

**Theorem 6.2.** If \( \omega \in \Omega \) is strongly palindromic, then \( H_\omega \) has no eigenvalues.

How can this theorem be applied to substitution hamiltonians? The authors of [15] provide a criterion which covers many interesting examples and which is easy to check for a given substitution. A primitive substitution \( \zeta \) belongs to class \( P \) if there is a palindrome \( p \) and for each \( a \in A \) a palindrome \( q_a \) such that \( \zeta(a) = pq_a \) for all \( a \in A \).

**Lemma 6.3.** If \( \Omega \) is coming from a class \( P \) substitution then it contains uncountably many strongly palindromic elements.

The class \( P \) contains many prominent substitutions, in particular the Fibonacci substitution. The interesting point is that eigenvalues are excluded not only for one element of \( \Omega \) which is the case for every trace map approach so far. Moreover, Theorem 6.2 also applies to Schrödinger operators with potential coming from so called circle maps (which provide the original quasiperiodic setting for the Fibonacci operator), compare [15].
7. Outlook and open problems. As we saw, spectral properties of one-dimensional quasi-crystals are accessible, as long as the self-similarities can be described by a primitive substitution. Of course, the mathematical approach presented is inherently restricted to the one-dimensional case because it is based on transfer matrix techniques. Hence, multidimensional quasi-crystals or at least multidimensional substitution Hamiltonians should be an interesting object to study. This has been done so far in a number of papers but many questions remain open. But still in the one-dimensional case there are enough problems left: of course, a general result about absence of eigenvalues would be interesting. Moreover, the occurrence of purely singular continuous spectra for many substitution Hamiltonians motivates the study of dynamical properties of these systems, some results in this direction are already announced [16].

References


Received 19.07.1996

David Damanik  
Fachbereich Mathematik  
J. W. Goethe Universität  
D-60054 Frankfurt  
e-mail: damanik@math.uni-frankfurt.de