Trace maps for arbitrary substitution sequences

1993 J. Phys. A: Math. Gen. 26 2437

(http://iopscience.iop.org/0305-4470/26/10/014)

J. Phys. A: Math. Gen. 26 (1993) 2437–2443. Printed in the UK

Trace maps for arbitrary substitution sequences

Yshai Avishai†§ and Daniel Berend‡||

† ISSP, University of Tokyo, Roppongi Minato-Ku, Tokyo 106, Japan and Department of Physics, Ben-Gurion University, Beer-Sheva 84105, Israel

‡ Department of Mathematics and Computer Science, Ben-Gurion University, Beer-Sheva 84105, Israel

Received 15 October 1992

Abstract. We prove the existence and show how to construct trace maps for products of 2×2 matrices generated by arbitrary substitution sequences. The dimension of the underlying space of our trace map is lower than the one suggested recently by other authors.

The discovery of quasi-crystals [1] and their one-dimensional modelling has led to a deep mathematical study of Schrödinger operators with an arbitrary deterministic potential sequence. On the way from periodic to random sequences, investigation has been focused on quasi-periodic systems such as the Fibonacci chain [2], having a discrete intensity measure, and on aperiodic systems such as the Thue–Morse sequence [3], with a singular continuous Fourier intensity measure.

Both the Fibonacci and the Thue–Morse sequences belong to a class of sequences generated by a substitution [4] (see below). Two other sequences of this kind are the circle sequence [5] (which is quasi-periodic) and the Rudin–Shapiro sequence [6] (which is aperiodic with a completely continuous Fourier intensity measure). The difference between the first and second pair of sequences is that the Fibonacci and Thue–Morse sequences are generated by substitutions defined on two letters while for the circle and Rudin–Shapiro sequences the alphabets are larger. As the number of letters increases, it is more and more difficult to obtain a workable trace map (see below), which serves as one of the central tools employed in the investigation of the spectrum of quasi-periodic or aperiodic structures. Trace maps were first derived by Kohmoto, Kadanoff and Tang [2]; see also Kohmoto and Oono [7]. Recently trace maps have also been used in the investigation of transport properties of several one-dimensional sequences [8]. Trace maps exist, and can be effectively constructed, for sequences generated by a substitution acting on two letters [9] (such as the Fibonacci and the Thue–Morse sequences for which the trace map is one-dimensional). Recently, Kolář and Nori [10] have shown that trace maps of higher dimensions do exist for substitution sequences containing more than two letters (e.g., for the circle and Rudin–Shapiro sequences).

In this work we address this problem and find trace maps for an arbitrary substitution sequence. Our trace maps have lower dimensionality than those of Kolář and Nori, which makes them more attractive for actual applications.

§ Research supported in part by the binational USA–Israel Science Foundation and by the Sherman Foundations for the UK–Israel science exchange program.

|| Research supported in part by a grant from the Israel Science and Technology Ministry.

Let Σ be a finite alphabet, say $\Sigma = \{1, 2, \ldots, r\}$, and let σ be a substitution on Σ, namely σ is a function from Σ to Σ^*, the set of all (finite) words over Σ:

$$\sigma(k) = \sigma_{k1}\sigma_{k2}\ldots\sigma_{kq_k} \qquad k = 1, 2, \ldots, r \qquad (\sigma_{ki} \in \Sigma,\ 1 \leqslant i \leqslant q_k). \quad (1)$$

We extend σ to a mapping from Σ^* to Σ^* by

$$\sigma(x_1x_2\ldots x_s) = \sigma(x_1)\sigma(x_2)\ldots\sigma(x_s) \qquad x_1, x_2, \ldots, x_s \in \Sigma. \quad (2)$$

Substitutions may serve as a means of defining sequences of matrices. Given r initial square matrices $A_{10}, A_{20}, \ldots, A_{r0}$ of the same size and a substitution σ on $\{1, 2, \ldots, r\}$, we define the sequences of matrices $\{A_{kn}\}_{n=0}^{\infty}$, $1 \leqslant k \leqslant r$, recursively by

$$A_{k,n+1} = A_{\sigma_{kq_k}n}\ldots A_{\sigma_{k2}n}A_{\sigma_{k1}n} \qquad 1 \leqslant k \leqslant r \qquad n = 0, 1, 2, \ldots. \quad (3)$$

As an example of the relevance of substitution sequences to physical systems, imagine a physical structure consisting of a one-dimensional sequence of atoms located at points x_n such that the distances $x_{n+1} - x_n = y(s_n)$ are in one-to-one correspondence with the terms of the given substitution sequence $\{s_n\}$. If each atom acts as a scattering centre whose potential at a distance x from the atom is $v(x)$, the motion of a particle in the field of this chain of atoms is governed by the Schrödinger equation

$$\left[-\frac{d^2}{dx^2} + \sum_n v(x - x_n)\right]\psi = E\psi. \quad (4)$$

Alternatively, put the atoms on the set of integers and let each atom have a site energy $V_n = V(s_n)$. Then the Schrödinger equation in the tight-binding approximation reads

$$-(\psi_{n+1} + \psi_{n-1}) + V_n\psi_n = E\psi_n. \quad (5)$$

The solution of (3) or (4) is attempted through the transfer matrix approach. It is then immediately found that certain products of transfer matrices satisfy the same recursion relations as in (2). The transfer matrices, in the plane wave representation, belong to the group

$$\mathrm{SU}(1,1) = \left\{\begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix} : \alpha, \beta \in \mathbb{C}, |\alpha|^2 - |\beta|^2 = 1\right\}. \quad (6)$$

Analysing the behaviour of sequences of matrices defined by a substitution is non-trivial. One approach, designed to obtain at least partial information, is to exploit the trace map. One tries to derive a formula which enables the traces of the matrices to be computed at each stage in terms of those of the preceding stage. An instance where this approach has been used [7] is for the Fibonacci substitution, defined by $r = 2, \sigma(1) = 12, \sigma(2) = 1$. Here we have two sequences of matrices $\{A_n\}_{n=0}^{\infty}$ and $\{B_n\}_{n=0}^{\infty}$ satisfying

$$A_{n+1} = A_nB_n \qquad B_{n+1} = A_n. \quad (7)$$

The authors of [7] showed that, if A_n and B_n are 2×2 matrices of determinant 1, and we set

$$a_n = \operatorname{tr} A_n \qquad b_n = \operatorname{tr} B_n \quad (8)$$

then

$$a_{n+3} = a_{n+1}a_{n+2} - a_n \tag{9}$$

(and $b_{n+1} = a_n$).

More generally, Allouche and Peyrière [9] proved that for any substitution on a two-letter alphabet and corresponding sequences A_n and B_n of 2×2 matrices, one can effectively construct a polynomial mapping Φ of five-dimensional space into itself such that

$$(\operatorname{tr} A_{n+1}, \operatorname{tr} B_{n+1}, \operatorname{tr} A_{n+1}B_{n+1}, \det A_{n+1}, \det B_{n+1}) = \Phi(\operatorname{tr} A_n, \operatorname{tr} B_n, \operatorname{tr} A_nB_n, \det A_n, \det B_n). \tag{10}$$

In the special case where all matrices have determinant 1, the mapping is defined on three-dimensional space. This mapping was used for various specific substitutions, such as the Fibonacci substitution mentioned above, the Thue–Morse substitution

$$r = 2 \qquad \sigma(1) = 12 \qquad \sigma(2) = 21 \tag{11}$$

and the substitution generated by the period doubling sequence

$$r = 2 \qquad \sigma(1) = 12 \qquad \sigma(2) = 11. \tag{12}$$

In the last two cases the trace map reduces to two-dimensional space.

Kolář and Noriv [10] extended the above-mentioned construction to apply to substitutions on alphabets with an arbitrary number of letters. They considered in particular what is obtained for the following two cases:

(i) The substitution generated by the so-called circle sequence:

$$r = 3 \qquad \sigma(1) = 313 \qquad \sigma(2) = 13313 \qquad \sigma(3) = 12313. \tag{13}$$

For 2×2 matrices of determinant 1, a trace map on six-dimensional space was found.

(ii) The substitution generated by the the Rudin–Shapiro sequence:

$$r = 4 \qquad \sigma(1) = 13 \qquad \sigma(2) = 43 \qquad \sigma(3) = 12 \qquad \sigma(4) = 42 \tag{14}$$

where a trace map of eight-dimensional space was derived (again for 2×2 matrices of determinant 1).

The use of the trace map in all these cases is clear. Consider, for example, the case of a substitution on a two-letter alphabet, with matrices over $\mathbb{R}$. The map giving the entries of A_{n+1} and B_{n+1} in terms of those of A_n and B_n acts on $\mathbb{R}^8$. The trace map acts on $\mathbb{R}^5$, and should thus be more manageable. In the case of the determinants being 1, the first approach yields a mapping of some six-dimensional manifold in $\mathbb{R}^8$, the second—a mapping of $\mathbb{R}^3$. For the circle sequence we have a mapping of $\mathbb{R}^6$ instead of a nine-dimensional manifold in $\mathbb{R}^{12}$, for the Rudin–Shapiro sequence a mapping of $\mathbb{R}^7$ instead of a twelve-dimensional manifold in $\mathbb{R}^{16}$ (or, in fact, instead of a nine-dimensional manifold in $\mathbb{R}^{12}$, as will be explained later).

In this paper we construct an alternative trace map, again for substitutions on any size alphabet. Our trace map can also be effectively constructed, and the construction has been implemented using the Mathematica software developed by Glaubman [11]. It has an advantage relative to that of Kolář and Nori [10] in that it is lower-dimensional. For an alphabet of size r it acts on $(2^r + r - 1)$-dimensional space (on $(2^r - 1)$-space if all determinants are 1). The main ingredient in the proof is the following elementary lemma (which was certainly noticed by numerous matrix theorists).

Lemma. Let A and B be 2×2 matrices. Then

$$BA = [\mathrm{tr}(AB) - \mathrm{tr}(A)\,\mathrm{tr}(B)]I + \mathrm{tr}(A)B + \mathrm{tr}(B)A - AB \tag{15}$$

where I is the identity matrix.

Proof. The lemma can be easily proved by direct computation. It will be more instructive, however, to prove it using the Cayley–Hamilton theorem. Working with generic matrices we may assume A and B to be invertible. For an invertible 2×2 matrix M, the Cayley–Hamilton theorem gives

$$M + \det(M)M^{-1} = \mathrm{tr}(M)I.$$

Consequently

$$\begin{aligned} BA &= \mathrm{tr}(BA)I - \det(BA)(BA)^{-1} \\ &= \mathrm{tr}(BA)I - [\det(A)A^{-1}][\det(B)B^{-1}] \\ &= \mathrm{tr}(AB)I - [\mathrm{tr}(A)I - A][\mathrm{tr}(B)I - B] \\ &= [\mathrm{tr}(AB) - \mathrm{tr}(A)\,\mathrm{tr}(B)]I + \mathrm{tr}(B)A + \mathrm{tr}(A)B - AB. \end{aligned}$$

This proves the lemma. □

Theorem 1. Let $A_1, A_2, \ldots, A_r$ be 2×2 matrices. Define the following 2^r matrices

$$B_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r} = A_1^{\varepsilon_1} A_2^{\varepsilon_2} \ldots A_r^{\varepsilon_r} \qquad (\varepsilon_j = 0, 1, \quad 1 \leqslant j \leqslant r). \tag{16}$$

Then any monomial $A_{j_1} A_{j_2} \ldots A_{j_s}$ ($1 \leqslant j_i \leqslant r$ for $1 \leqslant i \leqslant s$) can be effectively written as a linear combination of the matrices $B_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r}$, namely

$$A_{j_1} A_{j_2} \ldots A_{j_s} = \sum_{\varepsilon_1=0}^{1} \sum_{\varepsilon_2=0}^{1} \cdots \sum_{\varepsilon_r=0}^{1} c_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r} B_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r} \tag{17}$$

where each coefficient $c_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r}$ is a polynomial in the traces

$$\mathrm{tr}\, B_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r} \qquad (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_r) \in \{0, 1\}^r$$

and the determinants

$$\det A_j \qquad 1 \leqslant j \leqslant r.$$

In fact, if the given monomial is not some $B_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r}$, then $j_i \geqslant j_{i+1}$ for some $1 \leqslant i \leqslant s-1$. If $j_i = j_{i+1}$ then replacing $A_{j_i}^2$ by $\mathrm{tr}(A_{j_i})A_{j_i} - \det(A_{j_i})I$, we write our monomial as a linear combination of two shorter monomials. If $j_i > j_{i+1}$, then, by the lemma, the given monomial can be written as a combination of four monomials. Out of these, three are shorter than the original monomial, and the fourth is the same as the original monomial with $A_{j_i} A_{j_{i+1}}$ replaced by $A_{j_{i+1}} A_{j_i}$. Clearly, within finitely many steps we get a linear combination as required.

The theorem obviously gives a trace map for any substitution. We define

$$B_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r,n} = A_{1n}^{\varepsilon_1} A_{2n}^{\varepsilon_2} \ldots A_{rn}^{\varepsilon_r} \qquad \varepsilon_j = 0, 1 \qquad 1 \leqslant j \leqslant r \qquad n \geqslant 0. \tag{18}$$

Each of the matrices $B_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r,n+1}$ is a monomial in the matrices A_{j_n}, and consequently a linear combination of the matrices $B_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r,n}$. In particular, $\mathrm{tr}\, B_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r,n+1}$ is a polynomial in the $2^r - 1$ traces $\mathrm{tr}\, B_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r}$, $(\varepsilon_1\varepsilon_2\ldots\varepsilon_r) \in \{0, 1\}^r - \{(0, 0, \ldots, 0)\}$ (note that $B_{0,0,\ldots,0,n} = I$) and the r determinants $\det A_j$. Since

$$\det A_{j,n+1} = \prod_{l=1}^{q_j} \det A_{\sigma_{jl},n} \tag{19}$$

we indeed get a trace map as asserted.

Remark 1. The trace map may often be reduced to a space of lower dimension than that given by the general construction. Thus, for matrices of determinant 1, we get a trace map of two- instead of three-dimensional space for the Thue–Morse substitution, and of six- instead of seven-dimensional space for the circle sequence [10].

The construction of the trace map is also possible for the case where in the recursion defining each matrix $A_{j,n+1}$ we have a monomial in the matrices A_{jn} and their inverses A_{jn}^{-1} (as follows again from the Cayley–Hamilton theorem). The only difference is that the trace map is a polynomial in the traces $\operatorname{tr} B_{\varepsilon_1\varepsilon_2\ldots\varepsilon_r}$, the determinants $\det A_j$ and, in addition, their inverses $\det A_j^{-1}$. (Thus, for determinants 1 there is absolutely no difference.) This shows that for the Rudin–Shapiro sequence (with determinants 1) one can obtain a trace map on seven-dimensional space, instead of fifteen-dimensional space, as should be expected in view of our results in the paper. In fact, the formulae

$$A_{n+1} = C_n A_n \qquad B_{n+1} = C_n D_n \qquad C_{n+1} = B_n A_n \qquad D_{n+1} = B_n D_n \tag{20}$$

easily imply

$$D_n = C_n A_n^{-1} B_n \tag{21}$$

so that effectively we have only three sequences of matrices. (Note that (21) holds only from the second place on; the initial matrices do not have to satisfy this relation.) We shall now present the trace map for this case. Denote

$$\begin{aligned} &a_n = \operatorname{tr} A_n \qquad b_n = \operatorname{tr} B_n \qquad c_n = \operatorname{tr} C_n \qquad d_n = \operatorname{tr} D_n \\ &e_n = \operatorname{tr} A_n C_n \qquad f_n = \operatorname{tr} A_n B_n \qquad g_n = \operatorname{tr} B_n C_n . \end{aligned} \tag{22}$$

Routine calculations yield

$$\begin{aligned} a_{n+1} &= e_n \\ b_{n+1} &= c_n d_n - a_n b_n + f_n \\ c_{n+1} &= f_n \\ d_{n+1} &= b_n d_n - a_n c_n + e_n \\ e_{n+1} &= e_n f_n - b_n c_n + g_n \\ f_{n+1} &= c_n d_n e_n - a_n^2 b_n c_n + a_n c_n f_n + b_n c_n - g_n \\ g_{n+1} &= a_n b_n c_n^2 f_n - b_n c_n e_n f_n - a_n^2 b_n^2 c_n^2 + a_n b_n^2 c_n e_n \\ &\quad + a_n b_n c_n d_n + b_n^2 c_n^2 - b_n c_n g_n - b_n^2 - c_n^2 + 2. \end{aligned} \tag{23}$$

Using slightly different quantities, as done by Kolář and Nori [10], one can obtain a somewhat simpler map. Namely, put

$$\begin{aligned} &a_n = \operatorname{tr} A_n \qquad b_n = \operatorname{tr} B_n \qquad c_n = \operatorname{tr} C_n \qquad d_n = \operatorname{tr} D_n \\ &e'_n = \operatorname{tr} A_n C_n^{-1} \qquad f'_n = \operatorname{tr} A_n B_n^{-1} \qquad g'_n = \operatorname{tr} B_n C_n^{-1} . \end{aligned} \tag{24}$$

Then

$$
\begin{aligned}
a_{n+1} &= a_n c_n - e'_n \\
b_{n+1} &= c_n d_n - f'_n \\
c_{n+1} &= a_n b_n - f'_n \\
d_{n+1} &= b_n d_n - e'_n \\
e'_{n+1} &= g'_n \\
f'_{n+1} &= e'_n f'_n - g'_n \\
g'_{n+1} &= -a_n b_n f'_n + b_n c_n e'_n f'_n - c_n d_n f'_n + f'^2_n - b_n c_n g'_n + b_n^2 + c_n^2 - 2.
\end{aligned}
\tag{25}
$$

(Compare with the eight-dimensional trace map of Kolář and Nori.)

Remark 2. The mapping describing the recursion of the matrices acts on $4r$-dimensional space whereas the trace map acts on $(2^r + r - 1)$-dimensional space. This, given in addition that the trace map captures only part of the information, would suggest that the trace map is completely useless for $r \geqslant 4$, unless the dimension of the underlying space can be significantly reduced. We do believe, however, that the trace map may be of importance for theoretical purposes even for large r.

Remark 3. One is often actually interested in the norms of the sequence of matrices. Our trace map can also serve for that purpose. In fact, for a 2×2 matrix $A = (a_{ij})$ over $\mathbb{C}$, put

$$
\|A\|^2 = \sum_{i,j=1}^{2} |a_{ij}|^2. \tag{21}
$$

Then $\|A\|^2 = \mathrm{tr}(AA^\dagger)$. Thus, if in addition to the matrices A_{jn} defined recursively by a substitution, we take the matrices $A^\dagger_{jn}$, we get a system arising from a substitution on a double-sized alphabet. The trace map will also capture the information on the norms of the matrices A_{jn}.

Let us conclude by stressing some possible advantages of the present construction compared with the one suggested by Kolář and Nori [10].

(a) We have a consistent way of representing the monomials in the matrices in terms of the old monomials, the coefficients being traces. In [10], the recursion relation deals only with traces, and the new traces are expressed in terms of the old ones. Thus, the present scheme enables one to study the matrices themselves and not just their traces.

(b) Most important, our trace map has much smaller dimension. In our case (assuming the determinants of the matrices equal 1) the trace map for a substitution sequence of r letters acts on a space of dimension $d_1 = 2^r - 1$ while the trace map suggested in [10] acts on a space whose dimension d_2 is obtained as a sum of terms, the last of whom is $(r-1)!$. For large r we then have $d_2 \approx d_1^{\log_2 r}$. Presently, substitution sequences with large r are not experimentally realizable. Hence, their importance is only academic. Yet, with the rapid advances in the fabrication of superlattices, real systems with increasing value of r may be grown in the near future.

We are grateful to J M Luck for valuable discussions on this subject.

References

[1] Shechtman D, Blech I, Gratias D and Chan J W 1984 *Phys. Rev. Lett.* **53** 1951

[2] Kohmoto M, Kadanoff L P and Tang C 1983 *Phys. Rev. Lett.* **50** 1870
Levine D and Steinhardt P J 1984 *Phys. Rev. Lett.* **53** 2477

[3] Thue A 1977 *Selected Mathematical Papers (Oslo, 1977)*
Morse M 1921 *Trans. Am. Math. Soc.* **22** 84

[4] Queffélec M 1987 *Substitution Dynamical Systems—Spectral Analysis (Lecture Notes in Mathematics 1294)* (Berlin: Springer)
Axel F, Allouche J P, Kleman M, Mendes-France M and Peyrière J 1986 *J. Physique. Coll.* **47** C3-181
Bellissard J 1986 *Statistical Mechanics and Field Theory: Mathematical Aspects, Proc. Int. Conf. (Groningen) (Lecture Notes in Physics 257)* (Berlin: Springer)
Riklund R, Severin M and Liu Y 1987 *Int. J. Mod. Phys.* B **1** 121
Axel F and Peyrière J 1988 *C. R. Acad. Sci. (Ser. II)* **306** 179; 1989 *J. Stat. Phys.* **57** 1013
Axel F 1989 *Quasicrystals and Incommensurate Structures in Condensed Matter: Mtg Third Int. on Quasicrystals* ed M Josè Yacaman, D Romeu, V Castano and A Gomes (Mexico: Vista Hermosa) p 243
Bellissard J 1989 *Number Theory and Physics (Proc. Winter School, Les Houches, France 1989) (Springer Proceedings in Physics 47)* vol 1, ed J M Luck, P Moussa and M Waldschmidt
Luck J M 1989 *Phys. Rev.* B **39** 5834
Bellissard J, Bovier A and Ghez J M 1991 *Commun. Math. Phys.* **135** 379
Axel F and Terauchi H 1991 *Phys. Rev. Lett.* **66** 2223

[5] Aubry S, Godrèche C and Luck J M 1988 *J. Stat. Phys.* **51** 1033

[6] Rudin W 1959 *Proc. Am. Math. Soc.* **10** 855
Shapiro B 1951 *Master's Thesis* MIT

[7] Kohmoto M and Oono Y 1984 *Phys. Lett.* **102A** 145

[8] Sutherland B and Kohmoto M 1987 *Phys. Rev.* B **36** 5877
Kollar J and Suto A 1986 *Phys. Lett.* **117A** 203
Holzer M 1988 *Phys. Rev.* B **38** 1709
Avishai Y and Berend D 1990 *Phys. Rev.* B **41** 5492; 1991 *Phys. Rev.* B **43** 6873; 1992 *Phys. Rev.* B **45** 2717

[9] Allouche J P and Peyrière J 1986 *C. R. Acad. Sci. (Ser.II)* **302** 1135

[10] Kolář M and Nori F 1990 *Phys. Rev.* B **42** 1062

[11] Glaubman D unpublished