Self-avoiding generating sequences for Fourier lattice designs

R. A. Bates, H. Maruri-Aguilar, E. Riccomagno, R. Schwabe, and H. P. Wynn

Abstract. Good designs, that is sets of sampling points, for multidimensional Fourier regression are based on integer lattices whose positive integer generators $\{g_1, \ldots, g_d\}$ have special self-avoiding properties. These properties lead to consideration of the generalised Nyquist rate, or the smallest sample size to achieve the properties. The self-avoiding property can be converted to a statement about the existence of integer vectors g which do *not* satisfy a special set of linear equations. It follows that some solutions are derived from certain problems of intrinsic interest such as Sidon sets, the Thue-Morse sequence and constructions based on integers with prohibited digits, related to the Cantor set.

1. Introduction: self-avoiding sequences

There is a general class of number-theoretic problems, some special cases of which have considerable literatures. These arise while investigating a special problem in statistical theory namely the construction of so-called D-optimal designs for multivariate Fourier regression [**14**] [**3**]. It follows that certain optimal solutions are related to well-known sequences such as Sidon sequences, the Morse-Thue sequence and a number of interesting Cantor-like constructions. For a thorough discussion of such sequences see [**1**].

Consider an infinite set of ordered positive integers

$$\mathcal{G} = \{g_1 < g_2, \ldots\}$$

which we call *generators* . Define the vector of the first d generators: $g^{(d)} = (g_1, \ldots, g_d)^T$. (We shall often write g for $g^{(d)}$ where the dimension is subsumed.) Let $\mathcal{A} = \{\mathcal{A}_r,\ r = 1, 2 \ldots\}$ be an infinite set of integer matrices with the following properties.

(1) Each A_r is an $n_r \times d$ matrix, where n_r is increasing in r.
(2) For integers $s < r$, each A_s is nested in each A_r in that A_s is the minor of A_r comprising the first n_s rows and first s columns of A_r.

1991 *Mathematics Subject Classification.* Primary 62K05, 42B05; Secondary 68Q45.

Key words and phrases. Experimental design, Nyquist sampling, Fourier series, self-avoiding sequences, sum-free sets, Sidon sequences, Cantor set.

UK EPSRC grant GR/S63502/01.

UK EP/D048893/1 (MUCM project).

UK EPSRC grant GR/S63502/01 and EP/D048893/1 (MUCM project).

We shall usually require an following invariance property.

Definition 1.1. The matrix sequence $\mathcal{A}$ is called invariant if for each $A_d \in \mathcal{A}$ the set of rows is invariant under permutation of entries.

Definition 1.2. A d-vector of integers $g^{(d)} = (g_1, \ldots, g_d)^T$ is said to be self-avoiding with respect to an $n \times d$ matrix A_d if all entries of $A_d g^{(d)}$ are non-null.

Definition 1.3. A d-vector of integers, $g^{(d)} = (g_1, \ldots g_d)$, is called self-avoiding up to d with respect to a nested sequence of matrices $A_1, \ldots, A_d$ and starting point s if all entries of $A_r g^{(r)}$, are non-null, $r = s, \ldots, d$.

Definition 1.4. An infinite sequence of positive increasing integers, $\mathcal{G}$, is called self-avoiding with respect to an infinite nested sequence of matrices $\mathcal{A}$ and starting point s if all entries of $A_r g^{(r)}$ are non-null, $r = s, s+1, \ldots$.

Note that typically we take $g_1 = 1$ and starting point $s = 2$.

The canonical problem addressed in this paper is: given $\mathcal{A}$ and $g_1 = 1$ find a sequence with minimal g_d such that $g^{(d)}$ is self-avoiding with respect to A_d. If we solve this for all d we say we have a solution to the global problem. This has the advantage that we are always optimal but the disadvantage that we may need to solve a different optimisation problem for each d and the solutions may not be nested, in the sense of Definition 4.

Alternatively, we can find a way of generating a single sequence $\mathcal{G} = \{g_1, g_2, \ldots\}$ such that we do quite well, in terms of minimizing g_d, although for each d we may not be optimal. One way of doing this is to use a *greedy* algorithm. Assume we have a solution solution up to g_d, then choose g_{d+1} to be the smallest integer which satisfies the property that all entries of $A_{d+1}g^{(d+1)}$ are non-null. Given g_1 this leads to a unique sequence and we simply call it the *greedy* solution. Third, we can try to generate a single sequence using some special iterative generation method. As we shall see, the greedy method sometimes gives such a sequence, and even when it does not it may still yield a sequence of considerable intrinsic interest.

Example 1. The natural numbers. We simply require that no integer is equal to a previous integer (see Peano's postulates). Starting at $s = 2$, we have, excluding sign changes,

$$A_2 = [1, -1],\ A_3 = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}, \ldots, A_d = \begin{bmatrix} 1 & -1 & 0 & \ldots & 0 \\ 1 & 0 & -1 & \ldots & 0 \\ & & \ldots & & \\ 0 & \ldots & & 1 & -1 \end{bmatrix}$$

The optimal solution, the greedy algorithm and the iteration $g_1 = 1, g_{d+1} = g_d + 1$ all gives the natural numbers.

Example 2. Sidon sequences. In its simplest form a Sidon sequence is a set $\mathcal{G}$ of integers such that the sums $g_i + g_j,\ \ i \leq j,\ g_i, g_j \in \mathcal{G}$ are different. A Sidon set corresponds to our finite case in Definition 2. Thus $\{1, 2, 5, 7\}$ is a Sidon set; the set of all $g_i + g_j$ is

$$\{2, 3, 4, 6, 7, 8, 9, 10, 12, 14\}$$

Much of the large literature concerns the Erdös conjecture that $d - \sqrt{g_d}$ is bounded. For an extensive review and bibliography see [**12**] . In our notation a typical A_d has rows with (i) one 1 and one -1 (ii) two 1's and two -1's, (iii) two 1's and one

-2; a row which has one 2 and one -2 is reduced to case (i). The greedy version is sometimes called the Chaola-Mian sequence.

Example 3. Sum free sets. We require that for no triple i, j, k, all different, with $g_i, g_j, g_k \in \mathcal{G}$ is it the case that $g_i + g_j = g_k$. Then clearly our A matrices have rows with one 1 and two -1's. The Erdös-Sloane conjecture, which has been proved, is that the number of sum-free subsets $\mathcal{G} \subset \{1, \ldots, N\}$ is $O(2^{\frac{N}{2}})$. [**4**] [**9**]

This paper is divided follows. The next section introduces the self-avoidance conditions from Fourier, with some examples. Section 3 shows a main results which is the optimality of a Cantor-set type of construction and this construction is used to expand into a wider discussion of special sequences in Section 4. Exact solutions with an algebraic geometry flavour are discussed in Section 6 and similar ideas used to give a bound in Section 6.

2. Optimal design for Fourier regression

The papers [**14**] and previous papers, give some motivation from optimal experimental design (sampling) for multivariate Fourier regression.

The one dimensional Fourier model of order m is

$$E(Y(x)) = \theta_0 + \sqrt{2}\sum_{r=1}^{m} sin(2\pi r x)\theta_r + \sqrt{2}\sum_{r=1}^{m} cos(2\pi r x)\phi_r, \quad 0 \le x \le 1$$

Following [**14**] we write $F(d; m_1, \ldots, m_d; M)$ for the complete Fourier model in d variables, $x_1, \ldots, x_d$, with "marginal" models of order $m_1, \ldots, m_d$, respectively and all "interactions" terms up to order M are included:

$$E(Y(x_1, \ldots, x_n)$$

$$= \theta_0 + \sqrt{2}\sum_{l=1}^{M}\sum_{k_1<\ldots<k_l}\sum_{r_1=1}^{m_{k_1}}\cdots\sum_{r_l=1}^{m_{k_l}}\sum_{\beta\in B_l} \sin(2\pi(\alpha_1 r_1 x_{k_1} + \cdots + \alpha_l r_l x_{k_l}) \times \theta_{k_1,\ldots,k_l,r_1,\ldots,r_l,\alpha}$$

$$+\sqrt{2}\sum_{l=1}^{M}\sum_{k_1<\ldots,k_l}\sum_{r_1=1}^{m_{k_1}}\cdots\sum_{r_l=1}^{m_{k_l}}\sum_{\beta\in B_l} \cos(2\pi(\alpha_1 r_1 x_{k_1} + \cdots + \alpha_l r_l x_{k_l}) \times \phi_{k_1,\ldots,k_l,r_1,\ldots,r_l,\alpha},$$

where, $B_l = \{\beta \in [-1]^l,\ \beta_1 = 1\}$.

A Fourier lattice design in d dimensions, with a single d-dimensional generator $g = (g_1, \ldots, g_d)$, and sample size n is defined as the set

$$\left\{\left(\frac{jg_1 \pmod n}{n}, \ldots, \frac{jg_d \pmod n}{n}\right),\ j = 0, \ldots, n-1\right\}$$

We may take $g_1 = 1$, when g_1 and n are mutually prime. See [**11**], for study of such designs in relation to low-discrepancy problems in number theory and integration.

In the same way that, provided the sample size $n \ge 2m + 1$, an equally spaced design on $[-1, 1]$ is D optimum for the one dimensional model of order m, Fourier lattice designs are D-optimal for Fourier models for special choices of g and n. These optimal choices of g are exactly of self-avoidance type: we require certain linear combinations of the g_i to be zero. One way to think about this is as a non-aliasing property. Aliasing, in both experimental design and time series means a special type of non-confusion between terms in a linear regression model. Another evocative term is *resonance*. This occurs in areas such as, music, acoustics and radar when one frequency is a multiple of another frequency, or some sums of frequencies are equal to other summ. Thus, one could think of the self-avoidance conditions as a type of non-resonance requirement.

The following result gives the optimality conditions for the above multidimensional model.

THEOREM 2.1. *In the m-dimensional model $F(d; m_1, \ldots, m_d; M)$ the n-point lattice design generated by $g = (g_1, \ldots, g_d)$ is D-optimal up to the S-factor interactions, $(S \leq M)$ if and only if the members in the first S rows of the array of the following array*

$$\begin{array}{lll}
r_k g_k & r_k \in \mathcal{N}_k & k = 1, \ldots, d \\
r_{k_1} g_{k_1} + r_l g_l & r_k \in \mathcal{N}_k \text{ and } r_l \in \mathcal{N}_l & 1 \leq k < l \leq d \\
\vdots & \vdots & \vdots \\
r_{k_1} g_{k_1} + \cdots r_{k_m} g_{k_m} & r_{k_1} \in \mathcal{N}_{k_1}, \ldots, r_{k_m} \in \mathcal{N}_{k_m} & 1 \leq k_1 < \cdots < k_m \leq d \\
\vdots & & \\
r_{k_1} g_{k_1} + \cdots r_{k_M} g_{k_M} & r_{k_1} \in \mathcal{N}_{k_1}, \ldots, r_{k_M} \in \mathcal{N}_{k_M} & 1 \leq k_1 < \cdots < k_M \leq d
\end{array},$$

where $\mathcal{N}_k = \{-m_k, \ldots, -1, 1, \ldots, m_k\}$, are distinct in the cyclic group of n integer $\mathcal{G}_n$ and also (for $S < M$) distinct from the members in the last $M - S$ rows, in $\mathcal{G}_n$.

In practice, we ask that the integers be distinct over the positive integers. Then they are automatically distinct over the cyclic group $\mathcal{G}_n$ where $n = 2n_{max}+1$, where n_{max} is the largest integer in the array. This n could be described as a *generalized Nyquist rate* . That is to say a large enough sample size such that all possible aliasing arising from the problem at hand can be avoided. In one dimension the Nyquist frequency is $2m + 1$ for the m-th order Fourier model, simplified to $2m$ when the mean is zero [**8**].

We give two examples, both of which appeared as the results of greedy algorithms in [**14**] but which have since been found to have been already have the subject of some study by other authors.

Example 4. $F(d; 1, \ldots, 1; 2), S = 1$. Magic numbers. The conditions are that:

$$g_i,\ g_i \pm g_j,\ 1 \leq i < j \leq d$$

are all different. The solution given by greedy algorithm is the sequence

$$1, 3, 8, 18, 30, 43, 67, 90, 122, 161, 202, 260, \ldots.$$

In [**15**] they are are called "magic integers". They appear as A004210 in Sloane's on-line encyclopedia of integer sequences [**16**], where other references can be found.

Example 5. $F(d; 2, \ldots, 2; 1)$. The conditions are that all

$$g_i,\ 2g_j,\ \ 1 \leq i, j \leq d$$

are different. The greedy algorithm gives the sequence

$$1, 3, 4, 5, 7, 9, 11, 12, 13, 15, 16, 17, 19, 20, 21, 23, 25, 27, 28, 29, 31, 33, \ldots.$$

It can be shown that these have a precise description. If we expand in binary sequence we obtain

$$1, 11, 100, 101, 111, 1001, 1011, 1100, \ldots$$

showing that the parity (= 1 if odd, = 0 if even) of the number of ones alternates in d. This property characterizes the sequence (see A003159 in [**16**]). There are two other intriguing properties. If we write 1 in the location where an integer is included in the sequence and 0 otherwise, we obtain the sequence:

$$1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, \dots$$

This is one version of the so-called period-doubling sequence obtained by starting with 1 and expanding according to the expanding iteration $1 \to 10,\ 0 \to 11$ (A096263 in [**16**]). By taking the partial sums mod 2 of this binary sequence we have the Thue-Morse sequence whose iteration is:

$$t_0 = 0,\ t_{2n} = t_n,\ t_{2n+1} = 1 - t_n$$

We shall return to such constructions in Section 4.

3. A Cantor-like construction

We develop a greedy algorithm for the case $F(k, \dots, k; 2)$, which also has connection to the Cantor set. In the language of experimental design this is a "resolution IV" design:

(1) main effects, Fourier terms up to order k, are not aliased (confounded) with the constant term,
(2) main effects are not aliased with each other,
(3) main effects are not aliased with interactions: terms up to order k involving two factors

The conditions are

(i) $\pm r_i g_i \neq 0,\ r_i = 1, \dots k;$
(ii) $\pm r_i g_i \pm s_j g_j \neq 0,\ r_i, s_j = 1, \dots, k;\ i \neq j;$
(iii) $\pm r_i g_i \pm s_j g_j \pm t_l g_l \neq 0,\ r_i, s_j, t_l = 1, \dots, k;\ i, j, k$ all different.

Condition (i) comes form the requirement that no main effect is aliased with the constant term, (ii) that main effects are aliased with each other and (iii) that main effects are not aliased with any two-factor interaction.

Recall that the standard Cantor set is a subset of $[0, 1)$ in which middle thirds have been successively removed. A different type of Cantor set is where we remove the last third rather than the centre third. It is this type which concerns us. It is represented by the set of all reals which have base 3 expansions with no digit equal to 2. Reversing this we shall be interested in all integers which have no 2 in their base 3 representation.

$$1 \to 1,\ 3 \to 10,\ 4 \to 11,\ 9 \to 100, 10 \to 101 \dots \tag{3.1}$$

We generalize this as follows.

Definition 3.1. Define $C_{d,k+1}$ to be the d-th integer whose expansion in base $k + 1$ only has 0 or 1 as digits.

The following theorem generalizes the result for $k + 1 = 3$ given in [**14**], and is more succinctly stated here.

Theorem 3.2. *The sequence $g_d = 1 + 2kC_{d,k+1}$ satisfies condition (i), (ii) and (iii) above corresponding to the Fourier problem $F(d; k, \dots, k; 2)$.*

Proof. We check conditions (i),(ii) and (iii)
(i). This is obvious as $r_i g_i = 0$ only if $g_i = 0$, and we clearly only need consider the case $+r_i g_i$.
(ii). First, it is clear that $r_i g_i + s_j g_j > 0$. Then, as $r_i g_i = r_i(mod 2k)$, $r_i g_i - s_j g_j = r_i - s_j \pmod{2k} \neq 0$ for $r_i \neq s_j$; and for $r_i = s_j$ we have $r_i g_i - s_j g_j = r_i(g_i - g_j) \neq 0$, since $g_i \neq g_j$.
(iii). Here we have to consider two cases.

(a) If i, j, l are not mutually different we may assume $i = l$. Then

$$r_i g_i \pm s_j g_j \pm t_l g_l = r_i \pm t_l \pm s_j \pmod{2k} = 0 \ (\mathrm{mod} 2k),$$

if and only if $s_j = \pm r f_i \pm t_l$. In which case:

$$r_i G_i \pm s_j g_j \pm t_l g_l = s_j(\pm g_i \pm g_j) \neq 0,$$

since $i \neq j$.

(b) Now assume i, j, k are all different. Without loss of generality we may assume $r_i = \max(r_i, s_j, t_l)$. Then,

$$r_i g_i \pm s_j g_j \pm t_l g_l = r_i \pm s_j \pm t_l \pmod{2k}.$$

This is zero and therefore zero mod $2k$ only if $r_i - s_j - t_l = 0$, so we can restrict ourselves to this case. Writing $c_d = C_{d,k+1}$, $(d = i, j, l)$ we then have

$$\begin{aligned} r_i g_i - s_j g_j - t_l g_l &= r_i(1 + 2kc_i) - s_j(1 + 2kc_j) - t_l(1 + 2kc_l) \\ &= 2k(r_i c_i - (s_j c_j + t_l c_l)) \end{aligned}$$

The last expression is zero only when $r_i c_i = s_j c_j + t_l c_l$.

We next use the properties of the $C_{d,k+1}$ to complete the proof. Note first that in the expansion base $k + 1$
(i) $r_i c_i$ only has digits "0" and "r_i"
(ii) $s_j c_j + t_l c_l$ only has digits "0", "s_j", "t_l" and "r_i" (as $r_i = s_j + t_l$).
Now, in the condition $r_i c_i = s_j c_j + t_l c_l$, which we are want to test that $s_j c_j + t_l c_l$ has only digits "0" or "r_i". For the "0" case since $s_j + t_l = r_i \leq k < k + 1$ we both c_j and c_k must have digits "0" in the corresponding position. In the "r_i" case, since $s_j + t_l = r_i < k + 1$, we must have a digit "1" for c_j and c_l, in the corresponding position. Thus, in both cases we have $c_j = c_l$ and hence $g_j = g_l$, a contradiction, which completes the proof.

The rate of the growth of $g_d = 1 + 2kC_{d,k+1}$ is the rate of growth of $C_{d,k+1}$. Counting digits we see that $d_n \sim n^\gamma$, where $\gamma = \log 2/\log(k + 1)$, which can be identified with the fractal dimension of the generalised Cantor set obtained by deleting $k - 1$ of $k + 1$ sub-intervals at each stage.

4. Morphisms and related constructions

Previous examples point to a the relationship between the self-avoidance which is the main concern of the paper and certain construction which arise in areas such as finite automata and the theory of formal languages.

First, we revisit the Thue-Morse sequence (A010060 in [**16**]) which starts

$$0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, \ldots$$

It has been rediscovered several time, leading to several equivalent ways of defining it. A neat way is that if $a(n)$ is the sequence and we counted the number of ones in the binary expansion of n then $a(n)$ is the parity.

Next, let us return to the case of $F(d; 2, \dots, 2; k)$, of the last section and recall that, by Theorem 7 the generator sequence is

$$g_d = 1 + 4C_{d,3},$$

where $\{C_{d,3}\}$ is the sequence of integers whose ternary (base 3) expansion do not have the digit 2. We started to give some terms of this series and here is a longer sequence (including 0):

$$0, 1, 3, 4, 9, 10, 12, 13, 27, 28, 30, 31, 36, 37, 39, 40, 81, \dots$$

This sequence is A005836 in [**16**] and has a number of other defining properties.

(1) If t belongs to $\{C_{d,3}\}$ then do $3t$ and $3t+1$ and *nothing else* belongs to $\{C_{d,3}\}$.
(2) $C_{2d,3} = 3C_{d,3}$ and $C_{2d+1,3} = 3C_{d,3} + 1$.

We use the bracket notation to describe the first of these properties. Thus we say $\{C_{d,3}\}$ is S_F where $F = \{3t, 3t+1\}$. The second property points to the "morphism" property of the sequence. This property is also inherited by our sequence $\{g_d\}$. By substituting $C_{d,3} = \frac{1}{4}(g_d - 1)$ we have: the surprisingly simple iteration for our generators:

$$g_{2d} = 3g_d - 2, \;\; g_{2d+1} = 3g_d + 2$$

Let us also define the *characteristic* sequence $\chi(n)$ of an integer sequence S is the indicator of the position of its integers:

$$\chi(n) = \begin{cases} 1 & \text{if } n \in S \\ 0 & \text{otherwise} \end{cases}$$

Thus, for Example 5 (adding 0) we had that the period doubling sequence was the characteristic sequence of the solution.

The method of generating the period doubling sequence using the expansion mapping: $1 \to 01,\ 0 \to 11$, is a special case of a more general construction using a *morphism*. The construction is split into three parts. We use a morphism on an alphabet to obtain an infinite fixed point sequence, map to binary sequence and then use these as the characteristic sequences for our sequence $\mathcal{G} = \{g_d\}$. The following sequence is A003159 in [**16**] and for this and the above example we have relied on [**2**], and [**10**].

First, write down an expanding morphism on a set of letters, such as

$$a \to ab,\ b \to cb, c \to bb \tag{4.1}$$

Starting with one letter, such as a, we then generate an infinite sequence by expansion at each stage:

$$a\ ab,\ abcb,\ abcbbbcb, \dots$$

The final infinite sequence is then a fixed point of the morphism (2). If we take this infinite sequences and map the map $a \to 1,\ b \to 1,\ c \to 0$ we obtain the binary sequence

$$1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, \dots.$$

This is the characteristic function of the series

$$0, 1, 3, 4, 5, 7, 9, 11, 12, 13, 15, 16, 17, 19, 20, 21, 23, 25, \dots. \tag{4.2}$$

This is the required sequence A003159 in [**16**]. It has three other properties (as with $\{C_{d,3}\}$): (i) in terms of the bracket notation it is S_F where $F = \{2t+1, 4t, 4t+1\}$, (ii) there is an iteration of the characteristic function

$$\chi(2n+1) = 1,\ \chi(4n) = \chi(n),\ \ \chi(4n+2) = 0$$

and (iii) the sequence has a "prohibited digit representation", as for $\{C_{d,k}\}$, namely it is the sequence of those integers whose binary expansions do *not* end in an odd number of zeros.

We can alternatively characterize the sequence by an equivalent self avoidance condition: for $i \neq j$,

$$2^k g_i - g_j \neq 0, \text{ for all odd } k$$

The proof is as follows. The condition that the binary expansion of g_i does not end in an odd number of zeros (which we will call good) is equivalent to $g_i = c2^b$, where c is odd and b is even. Assume the condition, then $g_j = 2^k g_i = c2^{b+k}$. But since k is odd and b even, $b+k$ is odd, so g_j is not good and we have self avoidance. Now assume self avoidance, so that $2^k g_i = g_j = c2^s$, where c is odd and s is odd. But then $k \leq s$ so that $g_i = c2^{s-k}$. But since s and k are odd $s-k$ is even and g_i is good.

We have a nice interpretation which returns us to our earlier discussion of resonance. Considering $2g_i$ as an octave, in the musical sense, we can state the self avoidance condition heuristically as "no odd octaves".

5. Hyperplane arrangements and exact solution

For fixed d and matrix A_d, let a_i^T be the i-th row of A_d. Let us introduce the notion

$$L_i(g) = a_i^T g,\ i = 1, \ldots, n_d.$$

Then we can consider the set of all L_i as a *hyperplane arrangement*. Following the usual practice, the hyperplane arrangement is considered as the union of the linear varieties (hyperplanes)

$$\mathcal{L}_i = \{g : L_i(g) = 0\},\ i = 1, \ldots, n_d,$$

and we can also consider the partially ordered lattice of all intersection of any subset of the $\mathcal{L}_i$ for $r = 2, \ldots, n_d$. See [**17**] for a thorough review of hyperplane arrangements.

Define the grid: $H_d = \{1, \ldots, m\}^d$. Algebraically this is the set of all solutions to the simultaneous equations

$$\prod_{j=1}^{m} (x_i - j),\ i = 1, \ldots, d. \tag{5.1}$$

For fixed A_d, the generator vectors g for which all entries of $A_d g$ are non-null, with maximum $g_d \leq m$, if they exist, lie in the set complimentary set

$$\bigcup_{i=1}^{n_d} \mathcal{L}_i \setminus H_d$$

A minimal solution to the global problem occurs for an ordered $g = (g_1, \ldots, g_d)^T$ with smallest g_d such that this set is empty.

We introduce another quantity from the theory of hyperplane arrangements: for fixed A_d the *defining polynomial* is

$$L(g) = \prod_{i=1}^{n_d} L_i(x).$$

It is clear that $\{g : L(g) = 0\} = \bigcup_{i=1}^{n_d} \mathcal{L}_i$.

LEMMA 5.1. *Let $L(x)$ be the defining polynomial of A_d. The g for which all entries of $A_d g_d$ are non-null are the solutions to the system of equations*

$$\mathcal{S}(A_d) = \left\{ L(g)v - 1 = 0, \prod_{j=1}^{m} (g_i - j) = 0,\ i = 1, \ldots, d \right\}$$

where v is a "dummy" variable. Moreover each solution in g gives a unique solution in v.

Proof. The first part follows from the fact that equation $L(g)v - 1 = 0$, prevents any of the $L_i(x)$ being zero. Then, given any solution g we solve for v: $v = L(g)^{-1}$. Note that the use of the dummy variable v in this way is a type of *saturation*, which is used in algebraic geometry to eliminate zeros in certain settings.

Lemma 8 implies that the number of distinct solution in $g = (g_1, \ldots, g_d)^T$ to $\mathcal{S}(A_d)$ are also the number of distinct solutions in (g, v). Thus we need only count the solution to the $\mathcal{S}(A_d)$. In the terminology of algebraic geometry, either the system $\mathcal{S}(A_d)$ has no solutions or it forms a variety V which is a set of points and the corresponding ideal is an *ideal of points*, I. Elementary theory tells us that the number of points is the number of monomials is the dimension of the quotient ring $k[g_1, \ldots, g_d]/I$, for the real field k. This can be counted by the Hilbert series $h(s)$; namely $h(1)$ gives the number of solutions. This yields the following. See [**6**], [**7**] for introductions basic material and [**5**] one useful software suite.

LEMMA 5.2. *For fixed d, m, A_d, the number of solutions g for all entries of $A_d g$ are non-null, and hence to the system $S(A_d)$ is $h(1)$ where $h(s)$ is the Hilbert series of the ideal of points defined by $S(A_d)$. If, in addition, the invariance property holds and $g_1 = 1$, the number of ordered solutions is $h(1)/(d-1)!$.*

Example 5 (cont.). Consider Example 5, with $d = 5$ and $m = 8$ then

$$L(g_1, g_2, g_3, g_4, g_5) = \prod_{i<j}^{5} (g_i - g_j)(g_i - 2g_j)(g_j - 2g_i)$$

and

$$S(A_d) = \{Lv - 1 = 0,\ \Pi_{j=1}^{m}(g_i - j) = 0,\ i = 1, ..5\}$$

The smallest $g_d = m$ for which there is a solution is $m = 7$ and, under the restriction $g_1 = 1$, the Hilbert series (with respect to a total degree ordering) is computed as

$$h(s) = 1 + 5s + 12s^2 + 19s^3 + 11s^4.$$

Using the lemma there are $h(1)/4! = 2$ solutions. They are $\{1, 3, 4, 5, 7\}$ and $\{1, 4, 5, 6, 7\}$. The first of these is from the greedy solution to Example 5.

When $m = 8$ the Hilbert Series is (with $g_1 = 1$)

$$h(s) = 1 + 5s + 14s^2 + 27s^3 + 38s^4 + 11s^5,$$

and $h(1)/4! = 4$, giving the previous solutions and two more: $\{1,3,5,7,8\}$ and $\{1,5,6,7,8\}$. We can continue this with higher values of m for fixed d, to find the number of solutions, which is a problem of interest.

5.1. A bound. We use a version of the probabilistic method to establish an upper bound for the minimum g_d in the global case.

THEOREM 5.3. *Let A_d be a $n_d \times d$ matrix with the invariant property, of Definition 1. Then there is a non negative integer d-vector g for which all the entries of $A_d g^{(d)} d$ are non-null, whose largest component is less than or equal to $n_d + 1$.*

Proof. We use a little of the hyperplane notation, of the last section. Thus, by definition, for an integer d-vector g, $A_d \neq_* 0$ if and only if $L_i(x) \neq 0,\ i = 1, \ldots, n_d$. Fix m, let N_i be the number of solutions g of $L_i(g) = 0$ which lie in the grid H_d. As these lie on intersection of a $(d-1)$-dimensional hyperplane with H_d, it is clear that $N_i \leq m^{d-1}$.

But then

$$\left|\{i : a_i^T g = 0, \text{ for some } i = 1, \ldots, n_d;\ g \in H_d\}\right| \leq \sum_{i=1}^{n_d} N_i \leq n_d m^{d-1}.$$

But if the last quantity is less than $|H_d| = m^d$, we have at least one $g \in \bigcup_{i=1}^{n_d} \mathcal{L}_i \setminus H_d$ and this is a solution of our problem. This is certainly achieved when $n_d m^{d-1} < m^d$ or $m \geq n_d + 1$. Finally, the invariance condition allows us find an ordered solution.

Note that the bound for Example 5 is $m = 32$, whereas the smallest is $m = 7$. We have the corollary that if n_d is polynomial in d then the growth of the bound is polynomial in d. The proof can be recast as the probabilistic method: choosing rows a_i^T at random and using the Bonferroni bound. Recall that the bound in Theorem 1 is for the global case. As mentioned, we have no guarantee that the greedy rate will give the same rate.

6. Discussion

This paper touches on a number different fields, but, in summary, mainly the following: (i) the self avoidance problem itself, (ii) constructions related to missing digits in n-ary expansion (iii) morphism-plus-map constructions and constructions based on the complements of hyperplane arrangements. We have also aware of interest in some of sequences such as Sidon sequences, which have their own constructive methods. We are left with many unsolved problems but one somewhat obvious, but strong, piece of intuition: self avoidance is difficult and needs complex even chaos-like patterns. Experience with the examples of this paper prompts the somewhat loose conjecture that, given $\{A_d\}$, and under some conditions a good or greedy solution can be derived from a morphism or prohibited digits sequence. Finally, as we saw with the Cantor construction an advantage of the prohibited digit construction is that we may be able to find the asymptotic rate fairly easily and related to a fractal dimension, of some kind.

References

[1] Allouche, J.-P., Shallit, J. Automatic sequences. Cambridge Univertity Press, Cambridge, 2003.

[2] Allouche, J.-P., Shallit, J., Skordev, G. Self-generating sets, integers with missing blocks, and substitutions. Discrete Math. 292 (2005), 115

[3] Bates, R. A.; Riccomagno, E.; Schwabe, R.; Wynn, H. P. Lattices and dual lattices in optimal experimental design for Fourier models. Comput. Statist. Data Anal. 28 (1998), 283–296
[4] Cameron P.J., Erdos, P. On the number of sets of integers with various properties, Number theory (Banff, AB, 1988), 61-79, de Gruyter, Berlin, 1990.
[5] CoCoATeam: CoCoALib: a C++ library for doing Computations in Commutative Algebra, Available at http://cocoa.dima.unige.it
[6] Cox D, J. Little and OShea, D Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra, 3rd ed., Undergraduate Texts in Mathematics, Springer, New York, 2007
[7] Cox D., Little, J. and O'Shea, D. Using Algebraic Geometry, Volume 185 of Graduate Texts in Mathematics. Springer-Verlag, New York, 1998
[8] . Brillinger, D. R. Time series: data analysis and theory. Holden Day, San Francisco, 1981.
[9] Green B. The CameronErdo"s conjecture, Bulletin of the London Mathematical Society 36 (2004),769–778
[10] C. Kimberling, Affinely recursive sets and orderings of languages, Discrete Math. 274. (2004), 147159
[11] Niederreiter H. Random number generation and quasi-Monte carlo methods. SIAM, Philadeplphia, 1992.
[12] O'Bryant, K. A complete annotated bibliography of work related to Sidon sequences", Electronic Journal of Combinatorics 11 (2004). 39
[13] Shi P.; Fang K.-T.; Tsai C.-L. Optimal multi-criteria designs for Fourier regression models Journal of Statistical Planning and Inference, 96 (2001), 387–401
[14] Riccomagno, Eva; Schwabe, Rainer; Wynn, Henry P. Lattice-based D-optimum design for Fourier regression. Ann. Statist. 25 (1997), 2313–2327
[15] Silverton J. V. On the generation of 'magic integers. Acta Cryst. (1978). A34, 634
[16] Sloane N. J. A. On-line encyclopedia of interger sequences. http://www.research.att.com/ njas/sequences/
[17] Stanley R. P. An introduction to hyperplane arrangements. in: Geometric Combinatorics, Eds: E Miller E., Reiner V. aand Sturmfels, B. IAS/Park City Mathematics Series 2007; Vol: 13

Department of Statistics, London School of Economics, London WC2A 2AE, UK
E-mail address: R.Bates@lse.ac.uk

School of Mathematical Sciences, Queen Mary University, Mile End Road E1 4NS UK
E-mail address: H.Maruri-Aguilar@qmul.ac.uk

Università degli Studi di Genova, Dipartimento di Matematica, Via Dodecaneso 35, 16146 Genova Italy
E-mail address: riccomagno@dima.unige.it

Otto-von-Guericke-Universität Magdeburg, Fakultät für Mathematik, Institut für Mathematische Stochastik, Postfach 4120, 39016 Magdeburg Germany
E-mail address: rainer.schwabe@mathematik.uni-magdeburg.de

Department of Statistics, London School of Economics, London WC2A 2AE, UK
E-mail address: H.Wynn@lse.ac.uk