

Hilbert Functions in Associative Algebras

L.H. Smith

May 18, 2005

Abstract

These are notes on thoughts that I had on the hilbert function for free (non-commutative) associative agebras in Spring 2005.

1 Hilbert Functions

1.1 Hilbert functions. If $R = k[x_1, \dots, x_n]$ and I is a homogeneous ideal in R , then $S = R/I$ is a graded k -algebra. The *hilbert series* of S is the power series

$$H_S(t) = \sum_{d=0}^{\infty} \dim_k(S_d)t^d.$$

We will call $h_S(d) = \dim_k(S_d)$ the *hilbert function*. It is known that in this situation there is a polynomial $f_S(d)$, called the *hilbert polynomial*, that is equal to $h_S(d)$ for all d sufficiently large. The hilbert series of R itself is

$$H_R(t) = \sum_{d=0}^{\infty} \binom{d+n-1}{n-1} t^d.$$

It is always possible to compute the hilbert series, function and hilbert polynomial of a given ideal, and the general algorithm does this by finding the Gröbner basis of the ideal. In particular, the hilbert series of I is identical to the hilbert series of $\text{in}(I)$, the *initial ideal* generated by the highest terms of all polynomials in I determined by any well ordering of the terms. Of course, all ideals are finitely generated, as is the initial ideal itself.

1.2 Hilbert functions in free associative algebras. How is this theory extended to free associative algebras $R = k\{x_1, \dots, x_n\}$? We can, for example, use total degree in exactly the same way, so $\dim_k(R_d) = n^d$. For any “homogenous” ideal I there is a hilbert series. The hilbert series of R itself is

$$H_R(t) = \sum_{d=0}^{\infty} n^d t^d.$$

We cannot expect the limiting behavior in the hilbert series to be polynomial. But it is still the case that for any ideal, I and $\text{in}(I)$ have the same hilbert series, simply because $\text{in}(I)$ determines a basis in each degree for I_d .

1.3 Generators Q of quotient of a monomial ideal. Let R be a free k -algebra (k a field, for example) with free generators (or variables) x_1, \dots, x_n . Let $W = (w_1, \dots, w_m)$ be a finite set of (noncommuting) words in x_1, \dots, x_n and assume that $|w_i| \geq 1$. Also assume that W is reduced, that is there do not exist w_i, w_j , and words α, β such that $w_i = \alpha w_j \beta$ (if this occurs, w_i may be removed from W). If I is the two-sided ideal generated by W , then R/I is a k -algebra with a k basis given by

$$Q = \{ w \mid \forall \text{ words } \alpha, \beta \text{ and } w_i \in W, w \neq \alpha w_i \beta \}.$$

Let

$$Q_d = \{ w \in Q \mid |w| = d \}$$

then the general problem is to find the function $h_Q(d) = |Q_d|$.

2 Future Topics

2.1 Gröbner bases. Get the algorithm, I think from Mora, for computing Gröbner bases of two-sided ideals. Get a necessary and sufficient condition for a GB. Find a trivial though inefficient way to produce a GB, Given generators I , is a GB given by the union of all $I \cdot w$ where w is a word? Show that the GB of some principal ideals are not finite. Which ones, exactly?

Catalog ideals generated by one binomial, what are the GB? What are the principal generators in each degree? All principal ideals in degree 2 are GB, $(xy - yx)$ is a GB because $h_Q(d) = d + 1$ as expected. The polynomial $(xyx - yxy)$ is the “only” example in degree 3. Continue with one binomial and monomials. Continue with one arbitrary homogenous polynomial.

3 Finite State Formula.

3.1 Maximal overlap function. Given any word u , there exists a unique maximal length word v such that $u = \alpha v$ for some α , and $w_i = v\beta$ for some β and some $w_i \in W$, denote this word by $f(u)$. In words, v is the longest terminal sequence of u that is the initial sequence of some w_i . Notice that since W is assumed to be reduced, $u = \alpha w_i$ for some i if and only if $f(u) = w_i$. The values of $f(\cdot)$ are all initial segments of w_i , and therefore, there are only finitely many. Now define

$$S = \{ f(u) \mid f(u) \notin W \}$$

$$Q_{v,d} = \{ u \in Q_d \mid f(u) = v \}$$

Note that Q_d is a disjoint union of $Q_{v,d}$ for $v \in S$.

Claim: $\forall u \in Q_{v,d}$ and $\forall x_i f(ux_i) = f(vx_i)$.

Prove this.

3.2 Decomposition of Q . If $u \in Q_{v,d}$ and $d > 0$ then $u = u'x_i$ for some x_i and $u' \in Q_{d-1}$. Letting $v' = f(u')$, we have $u' \in Q_{v',d-1}$ and so $u \in Q_{v',d-1} \cdot x_i$.

Conversely, if $v' \in S$ and $f(v'x_i) = v$, then for any $u' \in Q_{v',d-1}$, $f(u'x_i) = v$ also. In particular, $u'x_i$ is not divisible by any w_i so $u'x_i \in Q_d$, and in fact $u'x_i \in Q_{v,d}$.

This shows that $Q_{v,d}$ is a union of $Q_{v',d-1} \cdot x_i$. The properties of words are such that this is a disjoint union,

$$Q_{v,d} = \bigcup_{\substack{v' \in S, x_i \\ f(v'x_i) = v}} Q_{v',d-1} \cdot x_i.$$

3.3 Enumeration formula for Q . Because of this, if $C_{v,d}$ is the number of words in $Q_{v,d}$ then

$$C_{v,d} = \sum_{\substack{v' \in S, x_i \\ f(v'x_i) = v}} C_{v',d-1}.$$

And $h_Q(d) = \sum_{v \in S} C_{v,d}$. Now express this compactly by first ordering $S = s_1, \dots, s_N$ with $s_1 = 1$. Define the $N \times N$ matrix M by

$$M_{ij} = |\{x_k \mid f(s_i x_k) = s_j\}|$$

and let C_d denote the $N \times 1$ column vector $(C_{s_1,d}, \dots, C_{s_N,d})$, then $C_d = MC_{d-1} = M^d C_0$ where $C_0 = (1, 0, \dots, 0)^T$. Finally, $h_Q(d) = \sum_i C_{i,d}$ so letting $\mathbf{1} = (1, \dots, 1)$ then

$$h_Q(d) = \mathbf{1} M^d C_0.$$

3.4 Example of enumeration formula. To illustrate, consider $W = \{x^2 y\}$ in $k\{x, y\}$. Then $S = \{1, x, x^2\}$ and

$$\begin{array}{ll} f(1 \cdot x) = x & f(1 \cdot y) = 1 \\ f(x \cdot x) = x^2 & f(x \cdot y) = 1 \\ f(x^2 \cdot x) = x^2 & f(x^2 \cdot y) = \text{not allowed.} \end{array}$$

Thus,

$$h_Q(d) = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}^d \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

It hardly bears mentioning that the growth of $h_Q(d)$ is asymptotically exponential. The largest eigenvalue of this matrix is $\lambda = (1 + \sqrt{5})/2 \approx 1.618\dots$, and $h_Q(d) = K\lambda^d + \dots$ and in fact the remaining terms are a constant and a negative exponential.

4 Enumeration

4.1 The shift graph. Given a set of monomials I_d of length d in n variables, what is $|I_{d+1}|$ and how does it relate to $|I_d|$. Define a directed graph on R_d

with edges for all pairs $(a, b) \in R_d$ in which $x_i a = b x_j$. If an edge (a, b) occurs in the graph, it can occur in only one way. However, the edge (b, a) may also occur. For any $a \in R_d$ there are precisely n edges starting at a . For let $a = a' x_j$, then for each distinct x_i , there is a distinct edge $(a, x_i a)$ since $x_i a = (x_i a') x_j$, and all edges arise in this way. It is clear that there is a Hamiltonian path, that is, a closed path visiting every vertex exactly once.

This graph needs a name, how about the *shift graph* in degree d . I also can't decide whether to use the terminology monomial and degree from polynomial theory, or to use word and length from group theory.

Now we have $|I_{d+1}| = 2n|I_d| - e$ where e is the number of edges contained in $|I_d|$. Therefore, for any d and $k = |I_d|$ there are bounds $e_d(k) \leq e \leq f_d(k)$.

4.2 Partitions of the shift graph. If I, J is a partition of R_d , let us say that the partition *cuts* an edge e if the head and tail of e are in different sets. Claim: the number of edges cut by any partition is even (when $n = 2$). (This can be seen by taking a Hamiltonian path. This path must traverse from words in I to words in J an even number of times. Which does not exactly prove the claim.) If the number of edges is $2k$, then $|I_{d+1}| = 2|I_d| + k$. So the problem of finding the minimum number of derived words reduces to finding the partition that cuts the minimum number of edges.

Lemma. If H is a subset of R_d which is connected by a simple closed path, or a union of words connected by disjoint closed paths, and if $a \notin H$, then there is an edge from a to some $b \notin H$. The proof is that if all the edges starting at a end in H , then those distinct words in H cannot all have predecessors in H unless a itself is in H . As a consequence, $R_d \setminus H$ is a union of disjoint closed paths.

4.3 Extending the shift graph by degree. The graphs for R_{d+1} can be derived from a graph for R_d by splitting nodes. Every word $a \in R_d$ becomes n words in R_{d+1} which are $a x_i$. Each pair of words a, b and a directed edge from a to b becomes n directed edges from words in the extension of a to words in the extension in b . Maybe this can be formulated to give an explicit description or construction of the graph of R_{d+1} .

4.4 Numeric representation of the edge matrix. Let the words of R_d be associated with n -ary numbers of d digits in the obvious way. To simplify the correspondence, assume the variables are x_0, \dots, x_{n-1} . Then a word $w = x_{i_1} \cdots x_{i_d}$ is associated with an integer, also denoted w , with $0 \leq w < n^d$,

$$w = \sum_{k=0}^{d-1} i_{d-k} n^k.$$

Consequently, in the graph of R_d there is an edge from a to b if and only if the integers are related by $b \equiv na + j \pmod{n^d}$ where $0 \leq j < n$. The $n^d \times n^d$ matrix C with a 1 in column a and row b (numbering columns and rows starting with 0)

if and only if there is an edge from a to b , has a simple form. For $0 \leq a < n^{d-1}$ the columns $a, a + n^{d-1}, a + 2n^{d-1}, \dots, a + (n-1)n^{d-1}$ are identical and have a 1 only in the rows b in the range $na \leq b < na + n$. If I is a $n^d \times 1$ column indicator vector (containing 0 or 1) then $I^T C I$ is equal to the number of edges contained within the corresponding subset of words in R_d . Therefore, from the previous formulation, the number of words in R_{d+1} that are derived from I is $I^T(2n - C)I$ where $2n$ is understood as a diagonal matrix.

4.5 Objective function. Suppose that we are seeking the extremes of $I^T C I$. Let $I = (v_0, \dots, v_{2^d-1})$, then

$$\begin{aligned} I^T C I &= \sum_{a,b=0}^{n^d-1} v_a c_{ba} v_b \\ &= \sum_{a=0}^{n^{d-1}-1} \left(\sum_{j=0}^{n-1} v_{a+jn^{d-1}} \right) \left(\sum_{k=0}^{n-1} v_{na+k} \right) \end{aligned}$$

For $n = 2$, C has 2^{d-1} null vectors and an eigenvector (all 1s) with value 2.

Note that $I^T C I$ can be optimized subject to the constraint that the entries in I are 0 or 1 and sum to k . But if Lagrange multipliers are used, it will be found that every such value of I is a critical point and the optimal value must still be found by searching. The values of the Lagrange multipliers have an intuitive definition (something like \pm the number of edges terminating in I), and the objective function can be expressed as a simple sum of the subset of them, but I do not see how any of this is useful.

4.6 Alternate description of shift graph. We can also define a graph on all words with an edge from word a to b whenever $b = ax_i$ for some variable x_i and an edge from b to c whenever $b = x_i c$ for some variable c . Note these edges are only between words of length differing by 1. The previous graph is obtained as a two step graph, that is there is an edge from a to b (of the same length) if and only if some c exists such that there is an edge from a to c and from c to b . What is interesting about this is that it works both ways, that both $|c| = |a| + 1$ and $|c| = |a| - 1$. The information I am getting at is that the condition that a and b generate common words of one greater length is equivalent to a and b being derived from a common word in one less length. What I am interested in, of course, is a condition on an ideal to have a minimal extension, similar to the lexicographic condition discovered by Macaulay [1], which applies also to its extension.

4.7 Numerical exploration of shift graph. It is possible to compute all subsets of R_d for $d \leq 4$ and $n = 2$ and to count the edges contained in the shift

Table 1: Minimum and maximum number of edges of the shift graph in degree 3 contained in subsets of vertices of size k .

k	min e	max e	# max
0	0	0	1
1	0	1	2
2	0	2	6
3	1	4	2
4	2	6	2
5	5	8	2
6	8	10	6
7	12	13	2
8	16	16	1

graph. Therefore, the minimum and maximum value of e for each size set can be determined. Results are shown in tables 1 and 2.

Note the symmetry in these tables, $\max_k - \min_k = \max_{2^d - k} - \min_{2^d - k}$. It is also striking that in degree 4 from $k = 3$ to $k = 4$ it is only possible to add 1 to the maximum.

4.8 Search for total order. Because of Macaulay's result, I speculate that there is a word order such that the largest words form a maximal edge partition and that the extension of a largest word set is again a largest word set in the next highest length. These two conjectures are practically independent. I can explore the possibility up to length 4, because there are necessary (and I hope to establish sufficient) conditions for an ordering on the semigroup to be extendable to a total order. More about that later. To see how I can go about the exploration, I need to search for maximal edge subsets of R_d of size k which nest neatly for successive k . That is, $I_1 \subset I_2 \cdots \subset I_s$ where $s = n^d$ and $|I_k| = k$ and each I_k is a maximal edge subset of R_d . It is feasible to generate all maximal edge subsets of R_d up to $d = 4$ for $n = 2$, and so it is feasible to search for complete nested sequences.

If a nested sequence could be found, say $I_k = I_{k-1} \cup \{i_k\}$, this defines an ordering on R_d given by $i_a < i_b \Leftrightarrow a < b$. This ordering can be tested to see if it satisfies the necessary (and sufficient?) conditions for being extendable to a total order, and whether it extends (and contracts) to R_{d+1} (and R_{d-1}) to subsets that are again maximal edge subsets. If all that turns out to be true for some nestings, the big question is, what is the order and how can it be defined intuitively.

Unfortunately, the evidence contradicts the goal because there are no nestings for $n = 2$ and $d = 4$ (although there are for $d = 3$). Explicitly, define the graph with vertices that are maximal edge subsets of R_d and there is an edge between I and J whenever they differ by one element only. For $d = 4$ there are only 74 maximal edge subsets (counting also the empty set and the whole

Table 2: Minimum and maximum number of edges of the shift graph in degree 4 contained in subsets of vertices of size k .

k	min e	max e	# max
0	0	0	1
1	0	1	2
2	0	2	6
3	0	4	2
4	0	5	12
5	0	7	4
6	0	9	2
7	2	11	6
8	4	13	4
9	6	15	6
10	8	17	2
11	12	19	4
12	16	21	12
13	20	24	2
14	24	26	6
15	28	29	2
16	32	32	1

graph). The graph edges are easily computed and the graph can be drawn to reveal pictorially that it is disconnected. (It is also possible to search for a nesting fairly efficiently, but since this fails the picture of the graph is better evidence.)

4.9 Alternative conjectures. What does this imply? Is it hopeless to continue? The problem is more fundamental than the existence of a total order on the monoids. There is no ordering in degree 4 of any kind. A tenacious mathematician might reason that the graph might be made connected if more vertices could be added to it. Thus, the imperfect picture might arise as an obstructed view of a perfect picture which has yet to be discovered. There seem to be several ways to attempt progress.

4.10 Extending the ring. The shift graph is at the root of the problem. An edge is joined between two monomials in R_d if they extend to the same element in R_{d+1} . In a general k -algebra, we would need to label the edge with the number of different elements that two elements can extend to. It should be possible to demonstrate this approach in the polynomial ring, where Macaulay gave the answer. The way to change the situation is to change the shift graph, and that means replacing R with a new algebraic structure. But what? Perhaps R can be extended, by adding new elements to each R_d . Or, perhaps we can

define a new kind of operation that generalizes multiplication.

4.11 Alternative orderings. As an alternative to modifying the ring, what we need is a condition that predicts when a set of words is a minimal generating set (previously I called them maximum edge sets). The order condition was that the set consists of all words less than or equal to a given word for some total order.

I can show that if any order could work, it must be a total order. Suppose that there were an arbitrary partial order and for all v , $\{w \mid w \leq v\}$ was a minimal generating set. If all sizes are obtained this way, then each $v \in R_d$ is associated with a k , so we can label them v_1, \dots, v_s (where $s = n^d$). But then it must follow that $\forall 1 \leq i < t \leq s$ that $v_i < v_t$, so the order is total.

Perhaps, we can seek a partial order which gives minimal generating sets but not for all k . This seems possible, but unsatisfying because it is incomplete.

4.12 Alternatives to orders. Rather than seeking an order, perhaps there is a geometric description. For example, if the elements of R_d are imbedded in a Euclidean space properly, it might happen that the minimal generating sets can be separated by a hyperplane. This might have the benefit of at least explaining why it is sometimes impossible to extend a set by a single element alone.

Or, perhaps we can simply use an automated search procedure to find an unconstrained logical condition that predicts when a set is minimal generating.

4.13 The second conjecture. Regardless of how the minimal generating sets (MGSs) are described, the second conjecture generalizing Macualay's result is that MGSs generate MGSs. But this is also false. In length 3 there are 6 MGSs of size 3 these are $0x03$, $0x11$, $0x81$, $0x24$, $0x88$, and $0xc0$. They all generate subsets of R_4 of size 6 containing fewer than 9 edges. But among subsets of R_4 there are 12 MGSs containing the maximal number of 9 edges. Also in length 3, there are only 2 MGSs of size 5, these are $0x37$ and $0xec$. The first one is

$$\{xxx, xxy, xyx, yxx, yxy\}$$

which generates

$$\{xxxx, xxxy, xxyx, xxyy, xyxx, xyxy, yxxx, yxxy, yxyx, yxyy, yyxx, yyxy\}$$

which has 12 elements and contains 20 edges in R_4 . The second one also generates a subset of R_4 with 20 edges. But there are 12 MGSs in R_4 that contain 21 edges.

So, not only is it the case that not all MGSs generate MGSs, but in some cases none of the MGSs of a given size generate MGSs. There is deficiency of both conjectures. They may be related in that their may exist one way of generalizing both conjectures that makes them both true.

5 Total orders on free semigroup

Here are some thoughts I had on total orders on the free semigroup, when I thought that was going to be important.

5.1 Degree respecting orders on free semigroup. This might be a paper that is an alternative approach for semigroups that parallels the paper on free groups. I am looking for particular total orders that respect the length, i.e. $|a| < |b| \Rightarrow a < b$. Then, suppose I have a prescribed ordering in length d . What are necessary and sufficient conditions on this ordering that it can be extended to a total order on the whole semigroup?

5.2 Multiplication maps. To simplify the remainder, let l_x/r_x denote mappings from $R_d \rightarrow R_{d+1}$ which are given by multiplication on the left/right by x , where x is any of the free generators. Then let M denote the set of all these maps so I can talk about any $m \in M$ without needing to consider whether it is a left or right multiplication.

5.3 Necessary condition: compatability below. Suppose that $<$ is a total order prescribed on R_d and $d > 1$. One necessary condition is fairly easy to see. For each $m \in M$ there is an order induced on R_{d-1} that is given by $a < b \Leftrightarrow m(a) < m(b)$. The necessary condition is that this induced order is a total order that is the same for all m . Furthermore, the induced order itself must satisfy the same condition for its degree, and the condition can be tested all the way to degree $d = 2$ (all degree 1 orderings are compatible with a total order). I will call this condition *compatability below at degree d* .

5.4 Necessary condition: compatability above. Another necessary condition is also not too bad. There is an induced partial order on R_{d+1} given by $m(a) < m(b)$ for all $m \in M$ and $a < b$ in R_d . Any total order that extends the original order must also respect these induced orders, hence there can be no cycles. I will call this condition *compatability above at degree d* .

5.5 Conjecture for sufficient condition. My conjecture is that compatability below implies compatability above. If that is true, then compatability below is a sufficient condition for the order to be extendable to a total order. The reasoning would go like this. Look at the order induced in R_{d+1} . Since it has no cycles, there a total order on R_{d+1} can be constructed (easily, though choices may be required), that is compatible with the extended partial order. Since it is compatible, it is compatible below at degree $d + 1$. Therefore, the process can be repeated indefinitely. The union of all of these total orders in each degree is finally a total order on all of R .

How to prove that compatability below implies compatability above? Start by assuming otherwise. Then there exists a cycle of length $t \geq 1$ in the induced partial order that can be described by m_i in M and $a_i < b_i$ in R_d for $i = 1, \dots, t$

such that $m_i(b_i) = m_{i+1}(a_i)$ and at the end, $m_t(b_t) = m_1(a_1)$. Somehow, there should be a way to derive a contradiction in degree d or lower, but it might be technical.

5.6 A distant goal. It would be nice if it could be proved that there is an intuitive geometric way to define the order, for example a linear form of some kind, which is different at each degree? For example, given a linear form in degree d that defines an order that is compatible with some total order, there is a procedure to derive a linear form in all smaller degrees that gives the order, and a procedure for extending the form to length $d + 1$.

6 MG sets, continued.

6.1 Symmetry of minimal generating sets. Let $P \subset R_d$, and $Q = R_d \setminus P$. We will let $e(P)$ denote the number of edges in the shift graph contained in P , and $e(P, Q)$ the number of edges originating in P and terminating in Q . Then we have

$$e(P) + e(Q) + e(P, Q) + e(Q, P) = e(R_d) = 2n^d.$$

Now P originates $n|P|$ edges and Q originates $n|Q|$ therefore $e(P, Q) = n|P| - e(P)$. It was proven above that $e(P, Q) = e(Q, P)$ therefore $n|P| - e(P) = n|Q| - e(Q)$. And so

$$e(Q) = e(P) - n|P| + n|Q|$$

This demonstrates that if P has the largest (smallest) value of $e(P)$ among all subsets of R_d with size $|P|$ then Q also has the largest (smallest) value of $e(Q)$ among all subsets of R_d with size $|Q|$. In order to find all subsets of R_d with extremal values of e , it is only necessary to search sets P with $|P| \leq n^d/2$. I think this is important also because it illustrates that the values of \max_k satisfy $\max_{n^d-k} = \max_k + n^{d+1} - 2nk$. For example, in table 2, $\max_{16-k} = \max_k + 32 - 4k$, or $\max_1 5 = \max_1 + 28$, $\max_1 4 = \max_2 + 24$, $\max_1 3 = \max_3 + 20$, etc. This takes some of the mystery out of the successive differences. Note also that the same equations hold for \min_k and the symmetry follows by subtracting: $\min_{n^d-k} - \max_{n^d-k} = \max_k - \min_k$ which was observed earlier.

6.2 Automorphism symmetry of MG sets. Any permutation of the variables in the free associative algebra is an automorphism, and any minimal generating set P is carried into another minimal generating set of the same size. There might be something that could be derived from this, but for now one can see this as an automorphism (symmetry) of the shift graph and the subset graph of the MG sets.

6.3 Median sets. I am aiming for a relation between the subsets such that it will always be possible to find minimal generating sets as subsets of other minimal generating sets. For example, if one begins with a minimal generating set P in R_4 with $|P| = 8$, then one can find minimal generating sets of all

smaller sizes by removing or adding one element at a time. Let me call sets P with $|P| = \frac{1}{2}|R_d|$ a median set (for $n = 2$, for larger n a different definition may be needed, eg with $\frac{1}{n}$). If it were true for all d , starting with P of size 2^{d-1} , it would allow the value of \min_k to be determined for all k . I might then hope that all such P could be characterized or computed in some simple way. The median sets for $d = 4$ are

- 1) $xxxx \quad xxxy \quad xxyx \quad xyxx \quad yxxx \quad yxxy \quad xxyy \quad yyxx$
- 2) $xxxx \quad xxxy \quad xxyx \quad xyxx \quad yxxx \quad yxxy \quad xyxy \quad yxyx$
- 3) $yyyy \quad yyyy \quad yyxy \quad yxyy \quad xyyy \quad xyyx \quad yxyx \quad xyxy$
- 4) $yyyy \quad yyyy \quad yyxy \quad yxyy \quad xyyy \quad xyyx \quad yyxx \quad xxyy$

Note that 1 and 4 are variable permutations, as are 2 and 3. Also, 1 and 4 can be reduced to MG sets of all smaller sizes by removing one element at a time, while 2 and 3 fail to generate an MG set of size 6 but can be extended to MG sets of all larger sizes by adding one element at a time. It does make sense that if one set can be reduced (or extended) that any permutation can likewise be reduced (or extended). So the question is, what is the difference between sets 1W and 2? Well, actually, they are complements, and whenever one set can be reduced by one, its complement can be extended by one.

For MG sets P , the number of out edges $e(P, \overline{P})$ appears to be well behaved, starting from 0 and increasing steadily to a maximum at the median set. For $n = 2$, I conjecture that the maximum is $e(P, \overline{P}) = d - 1$.

6.4 Another automorphism. Reversing the variables of words permutes the vertices of the shift graph and reverses all the arrows. Consequently, MG sets are carried to MG sets of the same size. This can be seen, for example, in the size 2 MG sets $\{xxxx, xxxy\}$ and $\{xxxx, yxxx\}$, while the set $\{xyxy, yxyx\}$ is fixed by the operation. This raises the question of whether there are other set automorphisms of R_d that carry MG sets to MG sets. One then wonders if the automorphism group could be determined more easily, could it then be used to determine the MG sets.

6.5 Lexicographic order, again. Consider the ring homomorphism from the free associative algebra to the polynomial ring. It appears that the inverse image of the monomials in a lexicographic ideal are MG sets (though it is clear that not all MG sets or even all sizes of MG sets are obtained in this way). Nevertheless, if it can be proved true, then it gives at least some MG sets. I thought of this conjecture when I realized that the median sets (for $n = 2$ and $d = 4$) consist of all words that have 3 or 4 x s in them plus half of the words with 2 x s. Then I realized that one could easily write a formula for the number of edges between groups of words having the same number of x s. These, of course, are just the inverse images of the monomials in the polynomial ring.

For the remainder of this subsection, the ring R is the free associative algebra with variables x and y . From earlier, R_d denotes the words of length d . Let $R_d(x^k)$ denote the set of words of length d that contain exactly k x s. Also let

$R_d(x^k, x-), R_d(x^k, -y), R_d(x^k, x-y)$, etc, denote those words in $R_d(x^k)$ that begin with x , end with y , or begin with x and end with y , and their variants. The first important observation is that all of the vertices of $R_d(x^k)$ have edges (in the shift graph) that point to either $R_d(x^{k-1}), R_d(x^k)$, or $R_d(x^{k+1})$. Explicitly, we can say that each vertex in $R_d(x^k, x-)$ has one edge pointing into $R_d(x^{k-1}, -y)$ and one edge pointing into $R_d(x^k, -x)$. Similarly, each edge of $R_d(x^k, y-)$ has one edge pointing into $R_d(x^k, -y)$ and one edge pointing into $R_d(x^{k+1}, -x)$. We can summarize this with

$$\begin{aligned} e(R_d(x^k), R_d(x^{k-1})) &= \binom{d-1}{k-1} \\ e(R_d(x^k), R_d(x^k)) &= \binom{d-1}{k-1} + \binom{d-1}{k} = \binom{d}{k} = |R_d(x^k)| \\ e(R_d(x^k), R_d(x^{k+1})) &= \binom{d-1}{k} \end{aligned}$$

In the subgraph $R_d(x^k)$, therefore, every vertex has in-degree 1 and out-degree 1. We can also say that if $P_k = R_d(x^d) \cup \dots \cup R_d(x^k)$ then $e(P_k, P_k^c) = \binom{d-1}{k-1}$. As for the median set, if we construct it as containing all $R_d(x^k)$ for $k > d/2$ (and if d is even, also add one half of all $R_d(x^{d/2})$ in the right way) then we might be able to get

$$e(P_{med}, -P_{med}) = \begin{cases} \binom{d-1}{(d-1)/2} & \text{if } d \text{ is odd} \\ \frac{1}{2} \binom{d}{d/2} & \text{if } d \text{ is even} \end{cases}$$

This is my updated conjecture on the median MG set. The number of edges added to $e(P, -P)$ when d is even to the next value of d is equal to the $e(P, -P)$ itself. That is, for d odd, $\min e(P_{med,d}, -P_{med,d}) = 2 \min e(P_{med,d-1}, -P_{med,d-1})$. And likewise there is a multiplier for d even, which converges to 2 from below.

6.6 More on the median sets. For d even, define P to be all of $R_d(x^q)$ for $q > d/2$ plus all the words in $R_d(x^{d/2})$ which are the end points of edges originating in $R_d(x^{d/2+1})$, these are precisely $R_d(x^{d/2}, -y)$.

Now count the edges in $e(P, -P)$. If any word in $R_d(x^q)$ with $q > d/2$, then its edges are in P by construction. So consider the words in $P \cap R_d(x^{d/2})$. Each originates two edges, one ending in $R_d(x^{d/2})$ and one ending in either $R_d(x^{d/2+1})$ or $R_d(x^{d/2-1})$ (the latter are not in P). In fact, all of the $R_d(x^{d/2}, y-y)$ have both edges that land back in P while $R_d(x^{d/2}, x-y)$ have edges neither of which land in P . Consequently, $e(P, -P) = 2|R_d(x^{d/2}, x-y)|$. But $|R_d(x^{d/2}, x-y)| = \binom{d-2}{d/2-1}$. This proves that, for d even, if P is a median MG set then $e(P, -P) \leq 2 \binom{d-2}{d/2-1}$.

For d odd, take P to be $R_d(x^q)$ for $q > d/2$. As before, any word in $R_d(x^q)$ with $q > d/2 + 1$ has both edges in P . The only edges that do not terminate in P are those in $R_d(x^{(d+1)/2})$ that end in $R_d(x^{(d-1)/2})$. There is one of these

for each of $R_d(x^{(d+1)/2}, x-)$ and as was seen before there are $\binom{d-1}{(d-1)/2}$ of these. This proves that, for d odd, if P is a median MG set then $e(P, -P) \leq \binom{d-1}{(d-1)/2}$.

Note. The “graph partition problem” is well known, in general NP-complete. It is the problem of partitioning a graph’s vertices into equal subsets V_1, V_2 such that the sum of edge (weights) between the two sets (also called the *cut*) is as small as possible. The problem of finding MG sets might therefore also be called the problem of finding minimal k -partitions.

Another note. This section had what I original thought of as a median set construction with the conjectured cut. Rereading it, I discovered an error, and there is no point in trying to fix it. For $d = 6$ I found a median set that had a greater $e(P)$. For this, take all $R(x^q)$ for $q > 3$ and then take $xxxyyy, xxyxyy, xyxxyy, yxxxyy, yxyxyy, yxyxxy, yyxxxy, yxyyxx, yxyxyx, yyyxxx$. I get $e(P, -P)$ to be 9, the conjectured minimum was 10.

6.7 The action of the cyclic group. The cyclic group \mathbb{Z}_d acts on R_d by cyclically shifting the words to the left, with the left-most variable becoming the right-most-variable. This action follows the edges of the shift graph and it preserves the “shape” of the word, that is the number of times each variable occurs in the word, and for $n = 2$, these are the $R_d(x^q)$. It is generally true, therefore, that the subgraph of the shift graph with the same shape is a disjoint union of cycles whose lengths are divisors of d . The remaining shift graph (for $n = 2$) consists in edges between $R_d(x^q)$ and $R_d(x^{q\pm 1})$ whose cardinality have already been observed.

It might be possible to derive formula for the number of components of various sizes in $R_d(x^d)$ (and for other shapes for higher n). This might give an interesting identity for the binomial (multinomial) coefficient, which was undoubtedly discovered by Gauss.

At any rate, the convenient construction of MG sets probably makes use of these cycles in some way. But there is certainly more structure to the shift graph. In fact, every two edges are connected by a path, even one that is part of a complete cycle. Perhaps there is a way of representing the graph in terms of all of those cycles such that the MG sets are related in some simple way. I think it might be interesting to draw the shift graph for each MG set to see just how the cycles relate.

6.8 More group actions. An automorphism of the shift graph is a permutation σ of R_d such that for every edge (a, b) , $(\sigma(a), \sigma(b))$ is also an edge in the shift graph. Now if σ is any permutation of the free generators x_1, \dots, x_n then applying it to the words induces an automorphism of the shift graph. The permutation of reversing the words might be called an anti-automorphism because it reverses the arrows in the graph. A group action is different. I would not require it to preserve edges, but rather to *follow* them. Specifically, let me define formally a graph permutation of the shift graph R_d is a permutation C of the vertices of R_d which satisfies

$$a \mapsto C(a) \Rightarrow (a, C(a)) \in E_d.$$

That is C follows the edges. Now define a group action on the shift graph to be any group of permutations of the vertices which is generated by graph permutations. Also, let CR_d denote the set of all edges $(a, C(a))$.

Now for some important examples. If σ is any permutation of the variables x_1, \dots, x_n then it induces a graph permutation, which will be denoted by L_σ defined by

$$L_\sigma : v_1 v_2 \cdots v_d \mapsto v_2 \cdots v_d \sigma(v_1).$$

In fact, I think it should be provable that all graph permutations are one of these. At any rate, one question is, what is the structure of the *full graph action group* (generated by all graph permutations)? What is the relationship between the subgroups of this full graph action group and the graph itself?

For each σ , the group generated by L_σ is a cyclic group of permutations of R_d . Therefore it is a trivial consequence that the orbits under its action are disjoint cycles whose lengths are divisors of $d \text{ ord}(\sigma)$. If σ is the identity, then the induced graph permutation preserves the number of each variable in a word, and the orbits are disjoint cycles of length dividing d . It is easy to visualize the shift graph when d is prime. Under the identity graph permutation, there are n cycles of length 1 (that is x_i^d) and the remaining words are in cycles of length d .

It is interesting to ask what the orbits of other cyclic group actions look like. If σ is a cyclic permutation of the variables (primitive, *ie* of order n), then the group generated by L_σ is cyclic and of order nd . To see this, note that the word $x_1^{d-1} x_2$ has an orbit of length nd . So it decomposes the graph into disjoint cycles. In fact, the edges are precisely

$$E_d = \bigcup_{i=1}^n L_{\sigma^i} R_d.$$

Therefore we see that the edges of the shift graph are a union of n sets of disjoint cycles. These cycles may be said to be transverse, *ie* crossing in some sense.

For $n = 2$, there is only one primitive cyclic permutation, and the the orbits are cycles of sizes dividing $2n$ and $E_d = L_e R_d \cup L_\sigma R_d$.

6.9 The orbits of graph actions Let σ denote any permutation of x_1, \dots, x_n , and let $D_{\sigma,d,k}$ denote the number of words in R_d that are fixed by L_σ^k . That is,

$$D_{\sigma,d,k} = |\{ w \in R_d \mid L_\sigma^k(w) = w \}|.$$

Given a computation of D will make it possible to compute the number of orbits of the action of L_σ of size exactly k . Let w be a word $v_0 \cdots v_{d-1}$ where each v_i is one of the free generators. Let $k = qd + r$ where $0 \leq r < d$. Then $L_\sigma(w) = w$ is equivalent to simultaneously satisfying d equations

$$\begin{array}{ccccccc} v_0 & \cdots & v_{d-r-1} & v_{d-r} & \cdots & v_{d-1} & \\ \parallel & \cdots & \parallel & \parallel & \cdots & \parallel & \\ \sigma^q v_r & \cdots & \sigma^q v_{d-1} & \sigma^{q+1} v_0 & \cdots & \sigma^{q+1} v_{r-1} & \end{array}$$

For convenience the indices of v are chosen here to correspond to the elements of the cyclic group \mathbb{Z}_d . With this interpretation, each equation is a relation between v_i and v_{i+r} , viewed as elements of \mathbb{Z}_d . Let $s = \gcd(d, k)$ and $p = d/s$. Then starting at any v_{i_0} there are p equations in a chain, where the variable appearing in the right hand side of one equation in the chain is the variable appearing in the left hand side of the next equation in the chain, and the variable appearing on the right hand side of the p th equation is again i_0 . There is a smallest index in the set of indices of the v_i that appear in such a chain, and it is less than s , so we may assume that each set of equations corresponds to an i_0 with $0 \leq i_0 < s$. The variables in this set of equations correspond to the p distinct indices $i_0, i_0 + r, i_0 + 2r, \dots, i_0 + (p-1)r$ where the arithmetic is understood modulo d .

Since $s|d$ and $s|r$, let $w = (d-r)/s$. Then for all $0 \leq i_0 < s$, and $0 \leq j < w$, $i_0 + js \in \{0, \dots, d-r-1\}$ and for $w \leq j < d$, $i_0 + js \in \{d-r, \dots, d-1\}$. Therefore, the systems of equations can be written explicitly as

$$\begin{aligned}
v_{i_0} &= \sigma^q v_{i_0+r} \\
v_{i_0+s} &= \sigma^q v_{i_0+s+r} \\
&\dots \\
v_{i_0+(w-1)s} &= \sigma^q v_{i_0+(w-1)s+r} \\
v_{i_0+ws} &= \sigma^{q+1} v_{i_0+ws+r} \\
&\dots \\
v_{i_0+(p-1)s} &= \sigma^{q+1} v_{i_0+(p-1)s+r}
\end{aligned}$$

Remember, all of the arithmetic on the indices takes place in \mathbb{Z}_d , so $i_0 + (p-1)s + r = i_0$.

For a given i_0 , the complete solution to this system of equations is determined from the value of v_{i_0} by working backwards through the chain starting at the last equation which gives $v_{i_0+(p-1)s}$, then finding the equation with $v_{i_0+(p-1)s}$ on the right hand side, and continuing. At each stage, the new variable is determined by applying σ^q or σ^{q+1} to the previous variable. The value of v_{i_0} gives a solution only if the final equation is satisfied which relates v_{i_0} to itself, and since all of the equations have been used, the exponent of σ in the last equation is the sum of all the exponents. This is $wq + (p-w)(q+1) = pq + p - w = k/s$. therefore the defining equation is

$$v_{i_0} = \sigma^{k/s} v_{i_0}.$$

And so the number of solutions to the system of equations for a given i_0 is the number of solutions to this last equation. And that number is the number of variables appearing in the cycle representation of σ in cycles whose length divides of k/s , which is independent of i_0 . So if this number is denoted by D_0 then

$$D_{\sigma, d, k} = D_0^s.$$

Table 3: The number (in parentheses) of points the given order under the action of L_σ on R_3 for the algebra generated by two noncommuting free variables.

d	k	$s = \gcd(d, k)$	k/s	num
3	1	1	1	0
3	2	1	2	2
3	3	3	1	0
3	4	1	4	2
3	5	1	5	0
3	6	3	2	8

If σ is the cyclic permutation of $1, \dots, n$, then

$$D_{\sigma^t, d, k} = \begin{cases} n^s & \text{if } tk/s \equiv 0 \pmod{n} \\ 0 & \text{otherwise.} \end{cases}$$

6.10 Calculation of orbits Take, for example, $n = 2$ with σ the interchange of x and y . Then $D_{\sigma, d, k} = 2^s$ where $s = \gcd(d, k)$ whenever $s \neq d$ and $pq + p - w$ is divisible by 2. The number of words of length 3 fixed by L_σ^k are given in Table 3. Note that in general, the graph necessarily repeats after $k = d \text{ord}(\sigma)$ which in this case is 6. The number of orbits of various orders can be determined from these numbers, using the fact that orbits are disjoint. There are potentially orbits of order 2, 4 and 6. Since $D_{\sigma, 3, 2} = 2$, there is just one orbit of order 2. Since $D_{\sigma, 3, 4} = 2$ also, there is no distinct orbit of order 4. And since $D_{\sigma, 3, 6} = 8$, 2 of these points are in the orbit of order 2 and the remaining points are in a single orbit of order 6.

The number of orbits of a given order k can be determined by counting the number of points of order k using the above formula and subtracting the number of points of smaller orders dividing k . This number of points is divisible by k (see next section), and the quotient is the number of distinct cycles. The Table 4 summarizes the number of orbits under L_{id} and L_σ .

6.11 An arithmetic formula. I haven't thought this through, so this is problem wrong. But something like it is correct.

The formula for the number of elements in cycles of order k under σ^t is

$$\sum_{i|k} \mu(k/i) D_{\sigma^t, d, i}$$

Therefore we have a kind of generalization of Fermat,

$$k | \sum_{i|k} \mu(k/i) D_{\sigma^t, d, i}$$

Table 4: The number (in parentheses) of cycles of the given order under the actions of L_{id} and L_σ on R_d for the algebra generated by two noncommuting free variables.

d	L_{id}	L_σ
3	1(2) 3(2)	2(1) 6(1)
4	1(2) 2(1) 4(3)	8(2)
5	1(2) 5(6)	2(1) 10(3)
6	1(2) 2(1) 3(2) 6(9)	4(1) 12(5)
7	1(2) 7(18)	2(1) 14(9)
8	1(2) 2(1) 4(3) 8(30)	16(16)
9	1(2) 3(2) 9(56)	2(1) 6(1) 18(28)
10	1(2) 2(1) 5(6) 10(99)	4(1) 20(51)

6.12 The shift group and automorphisms Consider the correspondence between the free variables to \mathbb{Z}_n , and R_d to \mathbb{Z}_n^d . Under this identification, σ corresponds to $v \mapsto v + s$ for some $s \in \mathbb{Z}_n$ and L_σ corresponds to the map

$$(v_1, \dots, v_d) \mapsto (v_2, \dots, v_d, v_d + s).$$

Let A denote the $d \times d$ matrix with

$$A_{ij} = \begin{cases} 1 & \text{if } i = j + 1 \text{ (modulo } n) \\ 0 & \text{otherwise} \end{cases}$$

where the entries may be interpreted as elements of \mathbb{Z}_n , and let $B(s) = (0, \dots, 0, s)$. Then under this correspondence, $L_\sigma(u) = Au + B(s)$. Therefore the shift group generated by all L_σ is isomorphic to the affine group

$$G_S = \{(A^k, B) \mid 0 \leq k < d \text{ and } B \in \mathbb{Z}_n\}$$

with the affine multiplication, which is simply the composition rule

$$(A^{k_1}, B_1)(A^{k_2}, B_2) = (A^{k_1+k_2}, A^{k_1}B_2 + B_1)$$

(with $k_1 + k_2$ reduced modulo d) and with identity $(A^0, 0)$. (I believe this is a semidirect product of the groups \mathbb{Z}_d and \mathbb{Z}_n^d). The group G_S is of order dn^d and is nonabelian. It acts naturally on \mathbb{Z}_n^d in a way isomorphic to the way the generators L_σ act on R_d . The generators L_σ correspond to the elements $g_s = (A, B(s))$ in G_S , denote these generators by G_L .

Now consider the graph automorphisms, which are permutations f of R_d that preserve edges

$$f : (u \xrightarrow{L_\sigma} v) \mapsto (f(u) \xrightarrow{L_\sigma} f(v))$$

for all σ . In the current language, these are permutations of \mathbb{Z}_n^d which commute with the group action. That is, f is a graph automorphism if and only if

$f(gv) = gf(v)$ for $g \in G_S$, or with $g = (A^k, B)$, $f(A^k v + B) = A^k f(v) + B$ for all k and all B . Taking $v = 0$ we get $f(B) = A^k f(0) + B$. But B can be identified with the elements of \mathbb{Z}_n^d , therefore $f(v) = A^k f(0) + v$. And since this definition cannot depend on k it must be the case that $A^k f(0) = A^0 f(0) = f(0)$, and this can only occur if $f(0)$ is a fixed point of A . The only fixed points of A are the vectors where all components are identical. This shows that all graph automorphisms are of the form $f(v)_i = v_i + s$ for some $s \in \mathbb{Z}_n$.

6.13 General graph automorphisms The most general kind of graph automorphism I can think of would simply eliminate the directions and require a 1-1 correspondence of edges. But I don't know how to formalize this. Instead, I will be content to formalize the two examples of general graph automorphisms, variable permutation and reversal.

If τ is any permutation of the variables, then it induces a permutation of the words of length d . Note the identity $\tau(L_\sigma(u)) = L_{\tau\sigma\tau^{-1}}(\tau(u))$. Thus an edge from u to $L_\sigma(u)$ corresponds to an edge from $\tau(u)$ to $\tau(L_\sigma(u))$ under $L_{\tau\sigma\tau^{-1}}$.

$$\tau : (u \xrightarrow{L_\sigma} v) \mapsto (\tau(u) \xrightarrow{L_{\tau\sigma\tau^{-1}}} \tau(v)).$$

If τ commutes with σ , then the result is a graph automorphism, and it was shown that all graph automorphisms are of this form. In fact, the only τ that commute with σ are cyclic permutations, i.e. $\tau = \sigma^s$ for some s (that is, $\tau(u) = u + s$).

Let r denote the reversal of variables in words of length d . Then note the identity $L_{\sigma^{-1}}(r(L_\sigma(u))) = r(u)$. Thus the edge from u to $L_\sigma(u)$ corresponds to an edge from $r(L_\sigma(u))$ to $r(u)$ under L_σ^{-1} :

$$r : (u \xrightarrow{L_\sigma} v) \mapsto (r(u) \xleftarrow{L_\sigma^{-1}} r(v)).$$

6.14 So long I am still struggling with computing MG-sets and coming up with a conjecture. There may be a binary quadratic programming algorithm, or branch and bound, which might be feasible up to $d = 6$ or 7 ; and that might give enough data to make a conjecture. However, there is a need for a theory. That is, a condition satisfied by MG sets that could be used to prove some structural and distributional properties. For example, it would be very good if it could be proven that MG sets are nested with small bounds on the "gaps." Working on this problem gave me some hope in my mathematical abilities, and I enjoyed thinking about occasionally. But now I leave it alone; I do not have time to work on it further.

References

- [1] Macaulay, F.S. (1927) Some properties of enumeration in the theory of modular systems. *Proceeds of the London Mathematical Society* 26, 531-555.