A simple construction for perfect factors in the de Bruijn graph

Chris Mitchell

c.mitchell@rhul.ac.uk

http://www.isg.rhul.ac.uk/~cjm

Agenda

1. de Bruijn sequences and arrays
2. Perfect factors
3. A new construction method
4. Further thoughts

De Bruijn sequences

- A (span *v*, *c*-ary) de Bruijn sequence is an infinite periodic sequence of symbols {0,1….,*c*-1} (for some $c>1$), with the property that every possible *v*-tuple of symbols occurs exactly once in a period.
- The period must clearly be equal to the number of *c*-ary *v*-tuples, i.e. c^v.

Examples

- [00011101] is a span 3, binary (2-ary) de Bruijn sequence (of length $2^3=8$).
- Here, as throughout, we write just one period of the bi-infinite sequence, and call it a **cycle** – we use square brackets for cycles.
- [001122021] is a span 2, 3-ary de Bruijn sequence (of length $3^2=9$).

Simple constructions I

- One very well-known method of generating binary de Bruijn sequences is a greedy algorithm, known as the ‘prefer one’ method.
- Start with the all zero *v*-tuple, and add one bit at a time, always adding a one **if possible**.
- E.g., for *v*=4:

[0000111101100101].

Simple constructions II

- Choose a prime power *q*.
- Take a maximum length *v*-stage shift register sequence over GF(*q*) (an m-sequence).
- This is a *q*-ary periodic sequence of period q^v-1, in which very *q*-ary *v*-tuple occurs except the all zero tuple.
- Take any one of the (*q*-1) all zero (*v*-1)-tuples in a period and insert a zero – this gives a span *v* *q*-ary de Bruijn sequence.

Pseudorandom sequences

- Equally, one can remove a zero from the unique all-zero *v*-tuple in any de Bruijn sequence to obtain a sequence which contains all possible *c*-ary *v*-tuples except the all-zero tuple.
- Such a sequence is known as a pseudorandom sequence.

Example

- [001122021] is a span 2, 3-ary de Bruijn sequence (of length $3^2=9$).
- [01122021] is a span 2, 3-ary pseudorandom sequence.
- [011220021] is another span 2, 3-ary de Bruijn sequence (of length $3^2=9$).

Existence

- De Bruijn (1946) and Good (1946) independently proved that de Bruijn sequences exist for every possible span *v* and every possible alphabet size *c*>1.
- They also gave an explicit formula for the number of such sequences for every *c* and *v*.
- It was subsequently discovered that their existence had first been established by Flye-Sainte Marie in 1894.
- Since the 1940s a significant number of different construction methods have been devised.
- Fredricksen (1982) gave an extremely helpful summary of construction techniques (since 1982 more techniques found).

de Bruijn-Good graph I

- This directed graph (which we write as $G(c,v)$) has vertices the c-ary v-tuples.
- Put an edge from vertex $(a_0,a_1,\ldots,a_{v-1})$ to vertex $(b_0,b_1,\ldots,b_{v-1})$ if and only if $a_{i+1}=b_i$ (for $0 \leq i \leq v-2$).
- A **Hamiltonian cycle** in $G(c,v)$ then corresponds to a c-ary span v de Bruijn sequence.

de Bruijn-Good graph II

- We can also label edges of $G(c,v)$ with c-ary $(v+1)$-tuples, i.e. so that the edge connecting $(a_0,a_1,\ldots,a_{v-1})$ to $(a_1,a_2,\ldots,a_{v-1},b)$ is labelled $(a_0,a_1,\ldots,a_{v-1},b)$.
- It is then not hard to see that an **Eulerian cycle** in $G(c,v)$ corresponds to a c-ary span $v+1$ de Bruijn sequence.
- Since the in-degree of every vertex is equal to the out-degree ($=c$) such an Eulerian cycle always exists – hence establishing the existence of de Bruijn sequences.

de Bruijn arrays (perfect maps)

- de Bruijn arrays are 2-dimensional analogues of de Bruijn sequences.
- An $(m,n;u,v)_c$-PM is a c-ary periodic array of period $m \times n$, in which every c-ary $u \times v$ sub-array occurs precisely once in a period (2-dimensional cycle).
- Hence, we must have $c^{uv}=mn$.
- Notion introduced by Reed and Stewart in 1962, who gave a $(4,4;2,2)_2$-PM.

Examples

- The Reed and Stewart example is as follows:

$$\begin{matrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{matrix}$$

- It is straightforward to verify that every 2×2 binary array occurs once when this is regarded as the recurring periodic pattern in an infinite array (i.e. a 2-dimensional cycle).

Necessary conditions

- The obvious necessary condition for the existence of a $(m,n;u,v)_c$-PM is $c^{uv}=mn$.
- For reasons it is simple to verify, we must also have:
  i. $u=m=1$ or $1 \leq u < m$;
  ii. $v=n=1$ or $1 \leq v < n$.
- These necessary conditions have been conjectured to be sufficient.

Existence

- The necessary conditions have been shown to be sufficient for the following cases:
  - c=2 (Paterson, 1994);
  - c a prime power (Paterson, 1996);
  - u=v=2 (Hurlbert, Mitchell and Paterson, 1996).

Applications

- A wide variety of applications have been proposed for de Bruijn sequences, including in:
  - cryptography;
  - position location.
- Position location/range funding applications have also been discussed for the two-dimensional arrays.

Agenda

1. de Bruijn sequences and arrays
2. Perfect factors
3. A new construction method
4. Further thoughts

Definition

- A perfect factor of a graph is a set of disjoint cycles of fixed length (*n*, say) which cover every edge.
- A perfect factor of the de Bruijn Graph $G(c,v\text{-}1)$ (which we call an $(n,v)_c$-PF) can be thought of as a set of c^v/n periodic *c*-ary sequences (of period *n*), for which every *c*-ary *v*-tuple occurs precisely once in a period of just one of the sequences.

Examples

- If $n=c^v$, then an $(n,v)_c$-PF is simply a span v, c-ary de Bruijn sequence.
- The set of two cycles: {[0,0,0,1], [1,1,1,0]} forms a $(4,3)_2$-PF.
- The set of three cycles: {[0,0,1], [1,1,2], [2,2,0]} forms a $(3,2)_3$-PF.
- The set of four cycles: {[0,0,3,3], [2,0,1,3], [1,1,2,2], [0,2,3,1]} forms a $(4,2)_4$-PF.

Necessary conditions

- We have the following trivial necessary conditions for the existence of an $(n,v)_c$-PF:
  1. $n|c^v$;
  2. $v=n=1$ or $1 \leq v < n$.
- These necessary conditions have been conjectured to be sufficient.

Applications

- Perfect Factors are simply a special case of perfect maps, since any ordering of the c^v/n cycles as the columns of a c-ary periodic array will form a $(c^v/n, n; 1, v)_c$-PM.
- [The converse is also trivially the case].
- Perfect Factors can also be used to help construct a much larger class of perfect maps (as we next see).

Etzion construction I

- Etzion (1988) showed how perfect factors can be used to construct perfect maps, generalising a construction of Ma (1984).
- We describe a special case of this construction proposed by Mitchell and Paterson in 1994.
- For simplicity we describe it for the binary case – however it works for arbitrary size alphabets.

Etzion construction II

- Let $C_0,C_1,\ldots,C_{n\text{-}1}$ be the cycles of an $(2^k,u)_2$-PF [and hence $k \leq u$].
- Let (r_i) be $2^{k(v-1)}$ repetitions of a (2^{u-k})-ary span v de Bruijn sequence [where $uv \geq 2k+1$].
- Let (s_i) be $2^{v(u-k)}$ repetitions of a 2^k-ary span $(v$-1) pseudorandom sequence for which the first v-2 elements are all zeros, preceded by $2^{v(u-k)}$ zeros.
- Let (w_i) be defined so that w_0=0, w_1=s_0, w_2=s_0+s_1, w_3=s_0+s_1+s_2, ...

Etzion construction III

- Then define a $2^k \times 2^{uv-k}$ array made up of columns from the perfect factor.
- Specifically, let the *i*th column consist of cycle C_{r_i} cyclically shifted by w_i places.
- This is a $(2^k,2^{uv-k};u,v)_2$-PM.
- Along with related constructions, this means that, if the perfect factor existence conjecture is positively resolved, the PM existence question will also be *mostly* resolved.

Existence I

- Etzion (1988) showed that $(2^k, v)_2$-PFs exist if $k \le v < 2^k$ [i.e. the necessary conditions are sufficient in c=2 case].
- Paterson (1994) showed the necessary conditions for an $(n,v)_c$-PF are sufficient if c is a prime power.
- Mitchell (1994) showed the necessary conditions are sufficient for all allowable triples $(n,v)_c$ as long as there exists a prime p such that $p^\alpha | n$ and $p^\alpha > v$.
- Mitchell and Paterson (1998) observed that, to completely resolve the existence question, it is only necessary to establish the existence of an $(n,v)_c$-PF for a 'square-free' c. This, in turn, means we only need to look at a finite number cases for each v.

Existence II

- The existence question has also been resolved for small values of *v* (the span).
- An $(n,v)_c$-PF always exists if:
  - v=2 (Mitchell, 1994);
  - $v \leq 4$ (Mitchell, 1995);
  - $v \leq 6$ (Mitchell and Paterson, 1998).
- It was also established that if a $(10,7)_{10}$-PF and a $(10,8)_{10}$-PF could be constructed, then the existence conjecture would be resolved for $v \leq 8$.

Agenda

1. de Bruijn sequences and arrays
2. Perfect factors
3. A new construction method
4. Further thoughts

Objective

- The main goal is to give a new construction method.
- We first describe a very simple construction which forms the basis of the new method.
- This method first appears in the 1998 Mitchell-Paterson paper, but I have a feeling it was known before.

A simple construction method

- Suppose n, c are integers greater than 1 such that $n|c^{n-1}$.
- Consider the set of c-ary cycles of length n whose elements sum to a value congruent to 1 modulo c – since $n|c^{n-1}$, these cycles have period exactly c.
- If we regard cyclic shifted sequences as equivalent, we obtain an $(n,n-1)_c$-PF.

Lempel homomorphism I

- We also need a homomorphism of the de Bruijn graph first given by Lempel (1970).
- The Lempel homomorphism D maps $G(c,v)$ to $G(c,v\text{-}1)$, and is defined by:

$$D(a_0,a_1,\ldots,a_{v-1}) = (a_1-a_0,a_2-a_1,\ldots,a_{v-1}-a_{v-2})$$

- D is a graph homomorphism – it is simple check that if there is an edge from $\boldsymbol{a}$ to $\boldsymbol{b}$, then there is an edge from $D(\boldsymbol{a})$ to $D(\boldsymbol{b})$.

Lempel homomorphism II

- This homomorphism turns out to be an incredibly useful tool in the study of de Bruijn sequences and related structures.
- It is analogous to differentiation, and has many related properties.

Lempel homomorphism III

- We can apply *D* to periodic sequences, as well as just to tuples.
- If the sequence (s_i) has period *u*, then $(D(s_i))$ will have period dividing *u*.
- Also, the mod *c* sum of *u* consecutive elements of $(D(s_i))$ is always zero.

The inverse homomorphism

- If $\boldsymbol{s}=[s_0,s_1,\ldots,s_{n-1}]$ is a cycle of weight w (reduced mod c) then we define the pre-image of $\boldsymbol{s}$, written $D^{-1}(\boldsymbol{s})$, to be the set of cycles:

$$\{\ [t,\ t+s_0,\ t+s_0+s_1,\ \ldots,\ t+(s_0+s_1+\ldots+s_{n-2}),\ t+w,\ t+w+s_0,\ \ldots\ ]\ \}.$$

- This set has size (w,c), and the cycles have period $nc/(w,c)$.
- Hence, if w mod c = 0, then $|D^{-1}(\boldsymbol{s})|=c$, and the cycles have period n.

The construction

- Suppose $c|n$ and c is odd.
- Suppose S is an $(n,n\text{-}1)_c$-PF constructed using the simple construction method outlined previously.
- Then $D(S)$ is an $(n,n\text{-}2)_c$-PF.

Example

- The set S of three cycles: {[0,0,1], [1,1,2], [2,2,0]} forms a $(3,2)_3$-PF (the sum of elements in each cycle is congruent to 1 mod 3).
- $D(S)$ = {[0,1,2]}, is a $(3,1)_3$-PF.

Justification I

Claim: If $\boldsymbol{a} = [a_0,a_1,\ldots,a_{n-1}] \in S$, then $D^{-1}(D(\boldsymbol{a})) \subseteq S$.

Proof: First note that the elements of $D^{-1}(D(\boldsymbol{a}))$ must have period n, since $D(\boldsymbol{a})$ must have weight 0.

If $\boldsymbol{b} = [b_0,b_1,\ldots,b_{n-1}] \in D^{-1}(D(\boldsymbol{a}))$, then for every i: $b_i=a_i+t$ for some t.

Hence:

$$b_0+b_1+\ldots+b_{n-1}= a_0+a_1+\ldots+a_{n-1}+nt \equiv 1 \pmod{c}$$

since $c|n$.

Justification II

- We next need to show that $|D^{-1}(D(\boldsymbol{a}))|=c$.
- Unfortunately, for $n=c=10$ this will not hold, since, for example,

  D([0000355558])=D([5555800003]).
- It will hold if we make the extra assumption that c is odd.

Justification III

- By the claim, the number of distinct cycles in $D(S)$ must be $|S|/n = c^{n-2}/n$.
- Hence the cycles in $D(S)$ contain a total of c^{n-2} $(n$-2)-tuples.
- Since every $(n$-1)-tuple occurs in a cycle in S, every $(n$-2)-tuple must occur in a cycle in $D(S)$.
- Hence, every $(n$-2)-tuple must occur in a unique cycle in $D(S)$.

Implication

- Unfortunately this construction does not give us a $(10,8)_{10}$-PF, since here *c* is <u>even</u>!

Agenda

1. de Bruijn sequences and arrays
2. Perfect factors
3. A new construction method
4. <u>Further thoughts</u>

Finishing the job

- Over the last 20 years we have assembled a powerful set of construction techniques for de Bruijn graph perfect factors.
- This in turn allows us to construct examples of perfect maps for ‘most’ parameter sets.
- However, there is the fear that from now on we will just be knocking off sporadic cases.

A general approach?

- It seems possible that the construction shown is just a special case of a much more general ‘simple’ construction technique.
- I am hopeful that this can be used to cover many more previously undecided parameter sets.
- Fundamentally, perfect factors are very numerous.