Periodicity of Degenerate Strings

Estéban Gabory[1], Eric Rivals[2], Michelle Sweering[1], Hilde Verbeek[1], and Pengfei Wang[2]

[1] Centrum Wiskunde & Informatica, Amsterdam, The Netherlands*
{esteban.gabory,michelle.sweering,hilde.verbeek}@cwi.nl
[2] LIRMM, Université Montpellier, CNRS, Montpellier, France**
{rivals,pengfei.wang}@limmr.fr

Abstract. The notion of periods is key in stringology, word combinatorics, and pattern matching algorithms. A string has period p if every two letters at distance p from each other are equal.
There has been a growing interest in more general models of sequences which can describe uncertainty. An important model of sequences with uncertainty are degenerate strings. A degenerate string is a string with "undetermined" symbols, which can denote arbitrary subsets of the alphabet Σ. Degenerate strings have been extensively used to describe uncertainty in DNA, RNA, and protein sequences using the IUPAC code (Biochemistry, 1970).
In this work, we extend the work of Blanchet-Sadri et al. (2010) to obtain the following results about the combinatorial aspects of periodicity for degenerate strings:

- We compare three natural generalizations of periodicity for degenerate strings, which we refer to as weak, medium and strong periodicity. We define the concept of total autocorrelations, which are quaternary vectors indicating these three notions of periodicity.
- We characterize the three families of period sets, as well as the family of total autocorrelations, for each alphabet size. In particular, we prove necessary conditions period sets should satisfy and, to prove sufficiency, we show how to construct a degenerate string which gives rise to particular period sets.
- For each notion of periodicity, we (asymptotically) count the number of period sets, by combining known techniques from partial words with recent results from number theory.
- Moreover, we show that all families of period sets, as well as the family of total autocorrelations, form lattices under a suitably defined partial ordering.
- We compute the population of weak, medium and strong period sets (i.e., the number of strings with that period set). We also compute the population of total autocorrelations.

Keywords: Periodicity, Degenerate string, Indeterminate string, Autocorrelation.

1 Introduction

Sequences of letters taken over an alphabet Σ, also called strings or words, are used to represent texts in natural languages, biomolecules such as DNA, RNA or proteins, or the sequence of states in dynamical systems. The notion of periodicity proves to be crucial for investigating word combinatorics [18], the properties of symbolic dynamical systems [19], or to design efficient pattern matching algorithms [9].

However, the classic notion of a string is insufficient to handle undetermination. Instead, sequences of sets of letters were considered and several definitions that generalize the classic notion of strings have been proposed such as partial strings and

* Address: P.O. Box 94079 - 1090 GB Amsterdam THE NETHERLANDS
** Address: 161 rue Ada - 34095 Montpellier cedex 5 FRANCE

Estéban Gabory, Eric Rivals, Michelle Sweering, Hilde Verbeek, Pengfei Wang: Periodicity of Degenerate Strings, pp. 42–56.
Proceedings of PSC 2023, Jan Holub and Jan Žďárek (Eds.), ISBN 978-80-01-07206-6

degenerate strings. In partial strings, the undetermined symbol $\diamond$ can represent any letter in Σ. In degenerate strings, we can have multiple different undetermined symbols, each representing a specified non-empty subset of Σ from which we must choose a letter. Degenerate strings thus generalize partial strings.

Regarding representations of biomolecules, reasons or causes of undetermination are multiple. First, undetermination appears in DNA/RNA sequences when sequencing machines fail to identify a precise nucleotide due to a noisy signal (which is frequent with the third generation of deep sequencing technologies, like Oxford Nanopore [21]. Second, undetermined symbols are used to represent binding sites (sequence regions at which biomolecules chemically bind to each other) at positions where alternative residues are observed. There exist databases of binding site representations using degenerate strings (JASPAR [5], HOCOMOCO [16]), which serve to identify new binding sites in genomes. Third, in the context of pangenomics, which investigates the genetic variations observed within a population, undetermined symbols serve to represent the variant nucleotides at a given genomic position in this population [28]. If only nucleotidic substitutions are considered, degenerate strings are adequate, but if multiple insertions/deletions need to be represented then elastic degenerate strings are preferred [13].

Related works. In classical finite strings, a period denotes the possibility of a word to self-overlap. In seminal articles Guibas and Odlyzko introduced the notion of period set of a finite word (and its binary representation the autocorrelation), proposed a characterization of it, investigated how the autocorrelation controls the probability of absence of a word in random texts, and extended it into correlation to study overlaps between pair of words. An alternative simpler proof of the "Fine and Wilf" theorem for period sets was given [10] and the set of period sets for words of length n and related combinatorics was investigated [24], while the asymptotic convergence on the number of period sets has recently been solved [25].

Our goal is to study combinatorics of period sets for string definitions allowing undetermined symbols. In a first step, Blanchet-Sadri et al conducted a combinatorial study in the case of partial strings [4,3] (similar to that on classical strings [23]), and investigated related algorithmic questions [2]. However, partial strings are inadequate to represent undetermination arising in biomolecules, while degenerate strings are. As degenerate strings generalize partial strings, we study combinatorics of period sets in the case of degenerate strings. In a related work, Iliopoulos and Radoszewski[14] showed that the weak period array of a degenerate string can be computed in $O(n\sqrt{n})$ and $O(n)$ space, while its strong period array cannot be computed in $O(n^{2-\epsilon}|\Sigma|^{O(1)})$ time if the Strong Exponential Time Hypothesis holds. Other algorithmic questions related to degenerate strings have also been investigated [6,12,11,26].

Contributions. For degenerate strings, three notions of period sets (weak, medium, and strong) are necessary and we characterize period sets for each, exhibiting necessary and sufficient conditions (Section 3). Then, we count the number of period sets for degenerate strings of length n for each notion and study its convergence using recent results from number theory (Section 4). We investigate the structure of the set of period sets in Section 5, and how many degenerate strings share a given period set (i.e., the population of a period set) extending the graph approach proposed in [3]. Finally, we outline some directions for future work (Section 7).

2 Preliminaries

A *classic string* $u = u[0\,..\,n-1] \in \Sigma^n$ of length n is a sequence of n letters over a non-empty finite *alphabet* Σ. For any $0 \leq i \leq j \leq n-1$, we denote the *substring* starting at position i and ending at position j with $u[i\,..\,j]$. In particular, $u[0\,..\,j]$ denotes a *prefix* of u and $u[i\,..\,n-1]$ a *suffix*. Throughout this paper, all our strings and vectors will be zero-indexed.

2.1 Degenerate strings

A *degenerate alphabet* Δ over Σ is a set of subsets of Σ, i.e., $\Delta \subseteq \mathcal{P}(\Sigma)$, where $\mathcal{P}(\Sigma)$ is the power set of Σ. We call the elements of a degenerate alphabet undetermined symbols, or *symbols* for short. A *degenerate string* $\hat{w} = \hat{w}[0\,..\,n-1] \in \Delta^n$ is a string of length n over the degenerate alphabet Δ. We define the *size* of $\hat{w}$ as the sum of the cardinalities of its symbols $\|\hat{w}\| = \sum_{i=0}^{n-1} |\hat{w}[i]|$.

Degenerate strings are used to model uncertainty. Undetermined symbols are used to denote all possible letters at a given position. This way, the degenerate string defines a language of words over the original alphabet Σ. Specifically, we define the *language* of a degenerate string $\hat{w}$ of length n over degenerate alphabet $\Delta \subseteq \mathcal{P}(\Sigma)$ as

$$\mathcal{L}(\hat{w}) = \{w \in \Sigma^n \mid \forall i \in \{0, \ldots, n-1\} \quad w[i] \in \hat{w}[i]\}\,.$$

Example 1. Let $\hat{w} = \left\{\begin{matrix}\mathtt{a}\\ \mathtt{b}\end{matrix}\right\} \cdot \left\{\begin{matrix}\mathtt{b}\\ \mathtt{c}\end{matrix}\right\} \cdot \{\mathtt{c}\}$. Note that $\hat{w}$ has length $|\hat{w}| = 3$, size $\|\hat{w}\| = 5$ and language $\mathcal{L}(\hat{w}) = \{\mathtt{abc}, \mathtt{acc}, \mathtt{bbc}, \mathtt{bcc}\}$.

A *hollow string* $\hat{w}$ is a degenerate string such that $\hat{w}[i] = \emptyset$ for at least one $i \in \{0, \ldots, n-1\}$, or equivalently a degenerate string such that $\mathcal{L}(\hat{w}) = \emptyset$. We say two degenerate strings $\hat{x}$ and $\hat{y}$ of length n over the same degenerate alphabet *match*, if for all $i \in \{0, \ldots, n-1\}$ the intersection $\hat{x}[i] \cap \hat{y}[i]$ is non-empty.

Sometimes, there are some restrictions on the degenerate alphabet Δ. In motif searching [22,27] for example, the k-motif which is a k-length degenerate string, consists of symbols such that the union of them is Σ and no symbol is a subset of another symbol. We will take $\Delta = \mathcal{P}(\Sigma) \setminus \emptyset$ unless stated otherwise, i.e., we have an undetermined symbol for every non-empty subset of Σ. This is the most general choice of Δ which excludes hollow strings. Hollow strings have an empty language, which is not very interesting when studying periodicity. Moreover, a nice consequence of excluding hollow strings is that now no two degenerate strings correspond to the same language.

2.2 Periodicity

Before we introduce the notion of periodicity in degenerate strings, we first recall its definition in the case of classic strings over the alphabet Σ. One such definition is as follows.

Definition 2 (Period of a string). *A string $u = u[0\,..\,n-1]$ has period $p \in \{0, 1, \ldots, n-1\}$ if and only if $u[0\,..\,n-p-1] = u[p\,..\,n-1]$, i.e., for all $0 \leq i \leq n-p-1$, we have $u[i] = u[i+p]$.*

There are several other equivalent definitions, e.g. one could require that $u[i] = u[j]$ whenever $i \equiv j \mod p$. Generalizing the notion of periodicity to degenerate strings is therefore not straightforward. Holub and Smyth introduced the concept of quantum and deterministic periods [11]. Blanchet-Sadri et al. call the same concepts weak and strong periods in the context of partial words [3]. We use the naming convention from Blanchet-Sadri et al. with the difference that we additionally define the concept of medium periodicity, which coincides with strong periodicity in the case of partial words but exhibits a different behaviour in the case of degenerate strings.

First, we recall the definition of weak periodicity.

Definition 3 (Weak period of a degenerate string). *A degenerate string* $\hat{w} = \hat{w}[0\,..\,n-1]$ *has weak period* $p \in \{0, 1, \ldots, n-1\}$ *if and only if* $\hat{w}[0\,..\,n-p-1]$ *matches* $\hat{w}[p\,..\,n-1]$*, i.e., for all* $0 \leq i \leq n-p-1$ *we have* $\hat{w}[i] \cap \hat{w}[i+p] \neq \emptyset$*.*

This is the most flexible type of periodicity, for which we want two strings in the language to overlap by $n-p$ letters. This type is most suitable when we use the degenerate string to model variations in a set of related strings.

Although periodicity p implies periodicity kp for classic strings, this is not the case for weak periods in degenerate strings (see Example 6). If we want to require this, we need a second stronger notion: medium periodicity.

Definition 4 (Medium period of a degenerate string). *A degenerate string* $\hat{w} = \hat{w}[0\,..\,n-1]$ *has medium period* $p \in \{0, 1, \ldots, n-1\}$ *if and only if for any* $0 \leq i, j \leq n-1$ *such that* $i \equiv j \pmod p$ *we have* $\hat{w}[i] \cap \hat{w}[j] \neq \emptyset$*.*

An equivalent definition is: a degenerate string $\hat{w}$ has medium period p if every multiple kp with $k \in \mathbb{N}$ is a weak period of $\hat{w}$. First notice that 0 is both medium period and weak period by definition.

Finally, we define strong periodicity.

Definition 5 (Strong period of a degenerate string). *A degenerate string* $\hat{w}$ *has strong period* p *if there exists a string* $w \in \mathcal{L}(\hat{w})$ *with period* p*.*

This is the most restrictive type of periodicity, where we require a word in the language to overlap itself. This type is most suitable when we use the degenerate string to model one specific string, of which letters are not precisely known.

Given a degenerate string $\hat{w}$, we denote its sets of weak, medium, and strong periods by $P^w(\hat{w})$, $P^m(\hat{w})$ and $P^s(\hat{w})$ respectively. From the definitions, we can easily see that $P^s \subseteq P^m \subseteq P^w$. We illustrate the difference between the different types of period sets with the following example.

Example 6. Let $\hat{w} = \begin{Bmatrix}\mathtt{a}\\\mathtt{b}\end{Bmatrix} \cdot \begin{Bmatrix}\mathtt{b}\\\mathtt{c}\end{Bmatrix} \cdot \begin{Bmatrix}\mathtt{b}\\\mathtt{c}\end{Bmatrix} \cdot \{\mathtt{c}\} \cdot \begin{Bmatrix}\mathtt{a}\\\mathtt{c}\end{Bmatrix}$. Then $P^w(\hat{w}) = \{0, 1, 2, 4\}$, $P^m(\hat{w}) = \{0, 2, 4\}$ and $P^s(\hat{w}) = \{0, 4\}$.

Finally, we denote the set of all possible weak, medium and strong period sets of degenerate strings of length n by Ω_n^w, Ω_n^m and Ω_n^s respectively.

2.3 Autocorrelations

One useful way to represent period sets is using autocorrelations, a concept introduced in 1981 by Guibas and Odlyzko [8]. The autocorrelation of a string $w \in \Sigma^n$ is the binary vector $s \in \{0, 1\}^n$ indicating its period set. We extend this definition by defining different autocorrelations for degenerate strings corresponding to different types of period sets.

Definition 7 (Autocorrelation of degenerate string). *For every degenerate string $\hat{w}$, its weak (resp. medium, resp. strong) autocorrelation is the binary vector $s \in \{0,1\}^n$ such that*

$$s[i] = \begin{cases} 1 & \text{if } i \text{ is a weak (resp. medium,} \\ & \text{resp. strong) period of } \hat{w} \\ 0 & \text{otherwise} \end{cases} \qquad \forall i \in \{0,\ldots,n-1\}.$$

We will denote the weak, medium and strong autocorrelations by $\hat{s}^w$, $\hat{s}^m$ and $\hat{s}^s$ respectively.

In [3], Blanchet-Sadri et al. take advantage of ternary vectors to simultaneously represent the weak and strong period sets of partial words. In our work, we introduce the concept of a total autocorrelation as a quaternary vector indicating these three notions of autocorrelations.

Definition 8 (Total autocorrelation of degenerate string). *For a degenerate string $\hat{w}$, its total autocorrelation is the sum of the weak, medium and strong autocorrelation $\hat{s} = \hat{s}^w + \hat{s}^m + \hat{s}^s$.*

We can equivalently define $\hat{s} \in \{0,1,2,3\}^n$ to be the vector such that

$$\hat{s}[i] = \begin{cases} 0 & \text{if } i \notin P^w & \text{(not a period)} \\ 1 & \text{if } i \in P^w \setminus P^m & \text{(weak period)} \\ 2 & \text{if } i \in P^m \setminus P^s & \text{(weak and medium period)} \\ 3 & \text{if } i \in P^s & \text{(weak, medium and strong period)} \end{cases}$$

for all $i \in \{0,\ldots,n-1\}$. To illustrate the weak, medium, strong and total autocorrelations, we review the degenerate string from example 6.

Example 9. Let $\hat{w} = \begin{Bmatrix} \mathtt{a} \\ \mathtt{b} \end{Bmatrix} \cdot \begin{Bmatrix} \mathtt{b} \\ \mathtt{c} \end{Bmatrix} \cdot \begin{Bmatrix} \mathtt{b} \\ \mathtt{c} \end{Bmatrix} \cdot \{\mathtt{c}\} \cdot \begin{Bmatrix} \mathtt{a} \\ \mathtt{c} \end{Bmatrix}$. Then $\hat{w}$ has weak autocorrelation $\hat{s}^w = 11101$, medium autocorrelation $\hat{s}^m = 10101$, strong autocorrelation $\hat{s}^s = 10001$ and total autocorrelation $\hat{s} = 31203$.

3 Characterization of total autocorrelations

In this section, we characterize the total (and hence also the weak, medium and strong) autocorrelation vectors of degenerate strings.

First, note that if the alphabet is unary, there exists a unique degenerate string of length n, which has total autocorrelation 3^n. Thus, we will henceforth assume that $|\Sigma| \geq 2$.

Theorem 10. *Let $P^s \subseteq P^m \subseteq P^w \subseteq \{0,\ldots,n-1\}$. Then P^w, P^m and P^s are respectively the weak, medium, and strong period sets of some non-hollow degenerate string $\hat{w}$ of length n if and only if*

A. $0 \in P^s$,
B. for all $p \in P^w$ *we have* $p \geq n/2 \implies p \in P^s$,
C. $p \in P^m$ *if and only if for all* $k \in \mathbb{N}$ *with* $kp \in \{0,\ldots,n-1\}$ *we have* $kp \in P^w$, *and*

D. $p \in P^s$ *if and only if for all* $k \in \mathbb{N}$ *with* $kp \in \{0, \ldots, n-1\}$ *we have* $kp \in P^s$.

Furthermore, these conditions are sufficient for any specific alphabet Σ *of cardinality at least 3. For a binary alphabet, we additionally require that* $P^m = P^s$.

Proof. We will first prove the necessity of these four properties. Let $\hat{w}$ be a degenerate string with weak, medium and strong period sets P^w, P^m and P^s respectively.

(I) Since $\hat{w}$ is not hollow, there exists a string $w \in \mathcal{L}(\hat{w})$. Since w has period 0, the degenerate string $\hat{w}$ has strong period 0.

(II) For every $p \in P^w$, there exist two strings $w_1, w_2 \in \mathcal{L}(\hat{w})$ such that $w_1[p\,..\,n-1] = w_2[0\,..\,n-1-p]$. Note that $w =: w_2[0\,..\,p-1]w_1[p\,..\,n-1] \in \mathcal{L}(\hat{w})$ as well. Moreover, since $p \geq n/2$, we have that $i \equiv j \mod p$ implies
- $i = j$ and hence $w[i] = w[j]$, or
- $j = i + p$ in which case $w[i] = w_2[i] = w_1[i+p] = w[i+p] = w[j]$, or
- $i = j + p$ and analogously $w[i] = w[j]$.

Thus, w has period p. Consequently, $\hat{w}$ has strong period p.

(III) This is the definition of medium periodicity.

(IV) Since $p \in P^s$, there exists $w \in \mathcal{L}(\hat{w})$ such that w has period p. If $kp < n$, then w also has period kp. Therefore kp is also a strong period of $\hat{w}$. Conversely, if kp is a strong period for all natural k such that $kp < n$, then trivially $1 \cdot p$ is a strong period as well.

To prove sufficiency, assume that $P^s \subseteq P^m \subseteq P^w \subseteq \{0, \ldots, n-1\}$ satisfy the four properties. We construct the degenerate string $\hat{w}$ such that

$$\hat{w}[i] = \begin{cases} \{\mathtt{a}, \mathtt{b}\} & \text{if } i = 0 \\ \{\mathtt{a}, \mathtt{c}\} & \text{if } i \in P^s \setminus \{0\} \\ \{\mathtt{b}, \mathtt{c}\} & \text{if } i \in P^w \setminus P^s \\ \{\mathtt{c}\} & \text{otherwise} \end{cases}$$

and verify that it has weak, medium and strong period sets P^w, P^m and P^s respectively.

- Note that every pair of sets intersects, except for $\{\mathtt{a}, \mathtt{b}\}$ and $\{\mathtt{c}\}$. Thus p is a weak period if and only if $\hat{w}[p] \neq \{\mathtt{c}\}$, which is indeed if and only if $p \in P^w$.
- The medium period set is defined by the weak period set by property (III). Thus, since $\hat{w}$ has the specified weak period set P^s, it also has the corresponding medium period set P^m.
- Note that for all $p \in P^s \setminus \{0\}$, the classic string $w \in \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}^n$ such that

$$w[i] = \begin{cases} \mathtt{a} & \text{if } p \mid i \\ \mathtt{c} & \text{otherwise} \end{cases}$$

is in $\mathcal{L}(\hat{w})$. Therefore $\hat{w}$ has strong period p. However, if $p \notin P^s$, then either $p \notin P^w$ (in which case p is not a weak period and thus not a strong period either) or there exists $k \in \mathbb{N}$ such that kp is a strong period with $n/2 \leq kp \leq n-1$ by property (II). It follows that

$$\hat{w}[0] \cap \hat{w}[p] \cap \hat{w}[kp] = \{\mathtt{a}, \mathtt{b}\} \cap \{\mathtt{b}, \mathtt{c}\} \cap \{\mathtt{a}, \mathtt{c}\} = \emptyset.$$

Therefore p is not a strong period.

We conclude that the four properties characterize the three period sets.

Note that the construction above uses an alphabet of size 3 and thus characterizes all possible total autocorrelations, even if we restrict to some specific alphabet Σ of cardinality at least 3. For binary alphabets, note that degenerate strings are the same as partial words, because they both have the same degenerate alphabet $\Delta = \{\{\mathtt{a}\}, \{\mathtt{b}\}, \{\mathtt{a}, \mathtt{b}\}\}$. Thus, every medium period is a strong period. In other words, the autocorrelation is in $\{0, 1, 3\}^n$. Conversely, any such autocorrelation is the autocorrelation of the binary degenerate string $\hat{w} \in \Delta^n$ such that

$$\hat{w}[i] = \begin{cases} \{\mathtt{a}\} & \text{if } i = 0 \\ \{\mathtt{a}, \mathtt{b}\} & \text{if } i \in P^w \setminus \{0\} \\ \{\mathtt{b}\} & \text{otherwise,} \end{cases}$$

because p is a weak period of $\hat{w}$ if and only if $p \in P^w$, and because the medium — and in the binary case also strong — periods are defined by the weak periods by property (III). □

4 Structure of autocorrelations

In this section, we take a closer look at the structure and number of weak, medium and strong autocorrelations.

4.1 Weak autocorrelations

We show that Ω_n^w, the set of autocorrelations of degenerate strings of length n with respect to weak periodicity, equals $\{1\}\{0,1\}^{n-1}$. This result holds irrespective of (non-unary) alphabet size.

Theorem 11. $\Omega_n^w = \{1\}\{0,1\}^{n-1}$

Proof. Let $s \in \{1\}\{0,1\}^{n-1}$. We construct a corresponding degenerate string over a binary alphabet $\{\mathtt{a}, \mathtt{b}\}$. We set $\hat{w}[0] = \{\mathtt{a}\}$, and for every $1 \leq i \leq n-1$, we set $\hat{w}[i] = \{\mathtt{b}\}$ if $s[i] = 0$ and $\left\{\begin{matrix}\mathtt{a}\\ \mathtt{b}\end{matrix}\right\}$ if $s[i] = 1$. It can easily be seen that s is the weak autocorrelation of $\hat{w}$. Note that any pair of symbols at position $i, j \geq 1$ in $\hat{w}$ has nonempty intersection $\{\mathtt{b}\}$. Therefore, we only need to observe that $\hat{w}[0]$ and $\hat{w}[p]$ match if and only if $s[p] = 1$. □

4.2 Medium and strong autocorrelations

Blanchet-Sadri et al. define $R(v)$ as the irreducible period set of partial word v and Φ_n to be the set of all irreducible period sets of partial words of length n [3]. They show that $R(v)$ is a primitive set, a set wherein no two numbers divide each other, and that any primitive subset of $\{1, \ldots, n-1\}$ is an irreducible period set. They also show that there is a one-to-one mapping between Φ_n and the number of period sets of partial words of length n. Consequently, to count the number of period sets, it is sufficient to count the number of primitive subsets of $\{1, \ldots, n-1\}$. In this section, we will similarly characterize the sets of medium and strong autocorrelations of degenerate strings.

Let us fix an integer interval $I = [0\,..\,n-1]$. Given a subset $P \subseteq I$, we write $\langle P \rangle = \{kp \in I \mid p \in P,\ k \in \mathbb{Z}_{\geq 0}\}$ and say that P *generates* $\langle P \rangle$. We say that P is *closed under multiplication* if $\langle P \rangle = P$. Note that this implies in particular that $0 \in P$.

This is a direct reformulation of Theorem 10 in the case of medium and strong autocorrelations:

Corollary 12. *The subsets of* $[0\,..\,n-1]$ *that are medium (resp. strong) period sets of a degenerate string* $\hat{w}$ *having length* n *over any fixed alphabet of cardinality at least* 2 *are exactly the multiplicative subsets of* $[0\,..\,n-1]$. *In particular, one has* $\Omega_n^m = \Omega_n^s$ *for any* $n \geq 1$.

We say that a set P of integers is *primitive* if it does not contain a pair $i \neq j$ such that i divides j. Equivalently, that means that $\langle P \rangle = \langle P' \rangle$ only if $P \subseteq P'$. Note that if P is a primitive set containing 0, then $P = \{0\}$.

Lemma 13. *Let* $I = [0\,..\,n-1]$. *Any set* $P \subseteq I$ *which is closed under multiplication contains a unique minimum set* P_{prim} *generating it, and this set is primitive.*

Therefore, primitive subsets in I *are in a* 1*-to-*1 *correspondence with multiplicative sets in* I.

Proof. The subset P_{prim} can be obtained by taking every pair $i \neq j$ with i dividing j and removing j. The order of removal does not affect the result by transitivity of the divisibility relation. Note that if $P \neq \{0\}$, then 0 will be removed from P as it is a multiple of every integer. The resulting set generates P and is primitive by construction. It is also the minimum generating set because if $\langle P_{\text{prim}} \rangle = P = \langle P' \rangle$ for some $P' \subseteq I$ then $P_{\text{prim}} \subseteq P'$ from the definition of a primitive set.

The reciprocal mappings $P \mapsto P_{\text{prim}}$ and $P \mapsto \langle P \rangle$ hence form a 1-to-1 correspondence. □

4.3 Counting the number of period sets

We have seen that weak period sets can be any subset of $\{0, \ldots, n-1\}$ containing 0. It follows that there are exactly $|\Omega_n^w| = 2^{n-1}$ weak period sets of strings of length n. Counting the number of medium and strong period sets is a lot more complex, but luckily we can rely on tools from the literature.

In [3], Blanchet-Sadri et al. provide upper and lower bounds on the number of autocorrelations of partial words of length n. They use a result by Erdős [7] to determine the logarithm of the number of primitive sets with elements smaller than n (and hence the number of autocorrelations of partial words of length n) up to a factor of two. However, recently there have been major developments concerning the number of primitive sets. Let $Q(n)$ be the number of primitive sets with largest element at most n. Angelo proved that $\ln(Q(n))/n$ converges to some constant α [1]. Liu, Pach and Palincza [17] and McNew [20] proved that α is effectively computable and computed upper and lower bounds on them.

Theorem 14 (Liu, Pach and Palincza [17], McNew [20]). *For any* $\epsilon > 0$, *we have*

$$Q(n) = \alpha^{n\left(1+O\left(\exp\left((-1+\epsilon)\sqrt{\log n \log\log n}\right)\right)\right)}.$$

The constant α *is effectively computable and* $1.5729 < \alpha < 1.5745$.

Since $|\Omega_n^m| = |\Omega_n^s| = Q(n-1)$, this implies the same asymptotic behaviour for the number of medium and strong period sets.

Corollary 15. *For any $\epsilon > 0$, we have*

$$|\Omega_n^m| = |\Omega_n^s| = \alpha^{n\left(1+O\left(\exp\left((-1+\epsilon)\sqrt{\log n \log\log n}\right)\right)\right)}.$$

The constant α is effectively computable and $1.5729 < \alpha < 1.5745$.

Proof. By Theorem 14, it follows directly that

$$|\Omega_n^m| = |\Omega_n^s| = \alpha^{(n-1)\left(1+O\left(\exp\left((-1+\epsilon)\sqrt{\log(n-1)\log\log(n-1)}\right)\right)\right)}.$$

Then one can notice that

$$\begin{aligned}
&(n-1)\left(1+O\left(\exp\left((-1+\epsilon)\sqrt{\log(n-1)\log\log(n-1)}\right)\right)\right)\\
&= n\left(1-\frac{1}{n}+\frac{n-1}{n}O\left(\exp\left((-1+\epsilon)\sqrt{\log n\log\log n}\right)\right)\right)\\
&= n\left(1+O\left(\exp\left((-1+\epsilon)\sqrt{\log n\log\log n}\right)\right)\right),
\end{aligned}$$

because $\frac{1}{n} = \exp(-\log n) = \exp\left(-\sqrt{\log^2 n}\right) = O\left(\exp\left(-\sqrt{\log n \log\log n}\right)\right)$. □

5 Lattice structure

Blanchet-Sadri et al. show that the sets of all binary and ternary autocorrelations of partial words of length n both form lattices under set inclusion of the corresponding period sets [3]. Moreover, they show these lattices satisfy the Jordan-Dedekind condition.

We investigate the structure of individual autocorrelations, as well as the total autocorrelation of degenerate strings. We show that Ω_n^w, Ω_n^m and Ω_n^s all follow lattice structure under set intersection and set union, and hence satisfy the Jordan-Dedekind condition. We also show the set of total autocorrelations is a lattice with respect to product order. Due to similarity with [3], we refer to Appendix A for the definitions of respectively the weak, medium, and strong autocorrelations, and for the proof of Theorem 16.

Theorem 16. *$(\Omega_n^w, \subseteq)$, $(\Omega_n^m, \subseteq)$ and $(\Omega_n^s, \subseteq)$ are lattices with respect to the inclusion order.*

In a poset (and hence also in a lattice), a *chain* is defined as a subset of totally ordered elements. The length of a chain is its cardinality minus one. The Jordan-Dedekind condition requires that all maximal chains between the same elements have equal length. If a lattice is distributive (i.e., $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ for all x, y, z in the lattice) and finite, then it satisfies the Jordan-Dedekind condition.

Since the meet and join of weak, medium and strong period sets correspond to set intersection and set union, we have the following corollary.

Corollary 17. *The lattices $(\Omega_n^w, \subseteq)$, $(\Omega_n^m, \subseteq)$ and $(\Omega_n^s, \subseteq)$ are all distributive and thus satisfy the Jordan-Dedekind condition.*

Let Ψ_n be the set of all total autocorrelations of length n. We will now show that Ψ_n is also a lattice with respect to product order (i.e., $u \leq v$ if and only if $u_i \leq v_i$ for all indices i) using the results for the individual families of period sets.

Theorem 18. *$(\Psi_n, \leq)$ is a lattice with respect to product order.*

Proof. To show that this is a lattice, we need to show that its meet ($\wedge$) and join operations ($\vee$) are well-defined. In this case, the meet will be the pointwise minimum of two total autocorrelations, while the join will be the minimum of all total autocorrelations greater than both. Formally,

$$u \wedge v = \min(u, v) \quad \text{and} \quad u \vee v = \bigwedge_{w \in \Psi_n \text{ s.t. } w \geq u,v} w.$$

Meet Let u and v be two total autocorrelations. Let P^w, P^m, P^s and Q^w, Q^m, Q^s be the weak, medium and strong period sets of u and v respectively. We define $R^w = P^w \cap Q^w$, $R^m = P^m \cap Q^m$ and $R^s = P^s \cap Q^s$. Note that $R^s \subseteq R^m \subseteq R^w \subseteq \{0, \ldots, n-1\}$ and
- $0 \in P^s \cap Q^s = R^s$,
- for all $p \in R^w = P^w \cap Q^w$ we have $p \geq n/2 \implies p \in P^s \cap Q^s = R^s$,
- $p \in R^m = P^m \cap Q^m$ if and only if for all $k \in \mathbb{N}$ with $kp \in \{0, \ldots, n-1\}$ we have $kp \in P^w \cap Q^w = R^w$, and
- $p \in R^s = P^s \cap Q^s$ if and only if for all $k \in \mathbb{N}$ with $kp \in \{0, \ldots, n-1\}$ we have $kp \in P^s \cap Q^s = R^s$.

Therefore there exists a degenerate string with weak, medium and strong period sets R^w, R^m and R^s respectively. Since we are taking intersections of the individual period sets, the corresponding total autocorrelation is the minimum of u and v.

Join Let u and v be two total autocorrelations. The join is the minimum of the autocorrelations greater than both u and v. Note that the minimum of all greater or equal autocorrelations is greater or equal than both u and v and not greater than any autocorrelation $w \geq u, v$. Observe that this join is well-defined since there is always at least one autocorrelation greater than or equal to both (namely 3^n) and that the join is an autocorrelation as well (because it is the meet of autocorrelations).

We conclude that $(\Psi_n, \leq)$ is a lattice. □

6 Population of autocorrelations

In this section we will give formulae to compute the population of autocorrelations of degenerate strings, in the case $\Delta = \mathcal{P}(\Sigma) \setminus \{\emptyset\}$. The population of an autocorrelation (resp. period set) is defined as the number of degenerate strings with this autocorrelation (resp. period set). We will follow the work of Blanchet-Sadri et al. [3], who compute the population number of partial words using graph theory. However, instead of looking at graph colourings, we look at independent sets to account for arbitrary sets of letters at each position of the degenerate string. We will first give the formulae for weak periods, and then explain how these can be adapted to find the population of medium periods. Finally, we discuss the case of strong periodicity and give an analogous hypergraph formulation to illustrate our difficulty in generalizing the result.

6.1 Weak and medium period sets

We are given a set $P \subseteq \{0, 1, \ldots, n-1\}$ and would like to compute how many degenerate strings there are over Σ with weak (resp. medium) period set P.

We define a graph on the set of positions $\{0, 1, \ldots, n-1\}$, with an edge connecting two vertices if and only if they differ by a period $p \in P$. We will first compute how many strings there are that have these periods (and possibly more periods). This is the number of ways we can assign subsets of Σ to the vertices such that

(a) no vertex is assigned the empty set, and
(b) the sets assigned to any two adjacent vertices have non-empty intersection.

We will first count the number of sets satisfying property (b) using the inclusion-exclusion principle. For each subgraph H we compute how many assignments there are where *all* pairs of adjacent vertices have *no* letter in common. For each letter there are $i(H)$ ways to assign it, where $i(H)$ is the number of independent sets in H. This gives $i(H)^{|\Sigma|}$ ways in total for the subgraph. There are $2^{(|V(G)|-|V(H)|)\cdot|\Sigma|}$ assignments for the rest of G. The number of assignments where every pair of adjacent positions has a letter in common — those satisfying property (b) — is thus

$$\sum_{H \subseteq G} (-1)^{|E(H)|} 2^{(|V(G)|-|V(H)|)\cdot|\Sigma|} i(H)^{|\Sigma|}.$$

Now, if every pair of adjacent vertices has a letter in common, all non-isolate vertices are assigned at least one letter. The isolate vertices are completely independent however, so we need to adjust for the chance of them being assigned the empty set, as this would result in a hollow string. Let $I(G)$ be the number of isolated vertices of G. By construction of the graph $I(G) = \max(2 \cdot p_{\min} - n, 0)$, where $p_{\min}$ is the smallest non-zero period in P (and n if it has no non-zero period). Removing the hollow strings we get

$$\left(\frac{2^{|\Sigma|}-1}{2^{|\Sigma|}}\right)^{I(G)} \cdot \sum_{H \subseteq G} (-1)^{|E(H)|} 2^{(|V(G)|-|V(H)|)\cdot|\Sigma|} i(H)^{|\Sigma|}$$

degenerate strings satisfying properties (a) and (b). This number contains all strings that have the given period set P as a subset of their period set. Thus to get the precise period, we must subtract bigger period sets using the inclusion-exclusion principle.

$$\sum_{P \subseteq Q \in \Omega_n} (-1)^{|Q|-|P|} \left(\frac{2^{|\Sigma|}-1}{2^{|\Sigma|}}\right)^{I(G_Q)} \cdot \sum_{H \subseteq G_Q} (-1)^{|E(H)|} 2^{(|V(G_Q)|-|V(H)|)\cdot|\Sigma|} i(H)^{|\Sigma|}$$

Here Ω_n is the set of all period sets and differs between the weak and medium cases.

6.2 Strong period sets

For strong periodicity, we can use the same technique. However, now we want that all positions with the same index modulo p have a letter in common. To model this, we can use the hypergraph $G = (V, E)$, where $V = \{0, \ldots, n-1\}$ and $E = \{\{j \in \{0, \ldots, n-1\} \mid j \equiv i \mod p\} \mid p \in P,\ i \in \{1, \ldots, p\}\}$.

We want to assign symbols to vertices such that for each hyperedge there exists a letter, which is in all symbols. Here things get more complex: if we want to use the inclusion-exclusion principle, we need to count the number of ways the constraints

on a certain set of hyperedges are violated. That is, for each such hyperedge and each letter, we do not want to assign the letter to all its vertices. Equivalently, the non-assigned vertices cover the hyperedges. Thus, if we define we define $i'(H)$ to be the number of vertex covers (also known as transversals) of H, then we can apply the same formula.

$$\sum_{P \subseteq Q \in \Omega_n^s} (-1)^{|Q|-|P|} \left(\frac{2^{|\Sigma|}-1}{2^{|\Sigma|}} \right)^{I(G_Q)} \sum_{H \subseteq G_Q} (-1)^{|E(H)|} 2^{(|V(G_Q)|-|V(H)|)\cdot|\Sigma|} i'(H)^{|\Sigma|}$$

Remark: Since $\Omega_n^m = \Omega_n^s$ for any $n \geq 1$, and some degenerate strings have a different medium and strong period sets, the population of a given period set should differ in medium and strong case. This is not the case for partial strings.

6.3 Total autocorrelations

To find the population of a total autocorrelation, we can use the same technique. Here, we choose the graph to be (V, E), where $V = \{0, \ldots, n-1\}$ and $E = E^w \cup E^m \cup E^s$, where E^w, E^m and E^s are the (hyper)edge sets corresponding to the weak, medium and strong period sets as defined above. The formula follows analogously.

$$\sum_{P \subseteq Q \in \Psi_n} (-1)^{|Q|-|P|} \left(\frac{2^{|\Sigma|}-1}{2^{|\Sigma|}} \right)^{I(G_Q)} \sum_{H \subseteq G_Q} (-1)^{|E(H)|} 2^{(|V(G_Q)|-|V(H)|)\cdot|\Sigma|} i'(H)^{|\Sigma|}$$

Remark: Note that these formulas are costly to compute. However, if we want to compute multiple populations, we can obtain slight speed ups using dynamic programming and memoization. For example, we can compute the number of independent sets $i(H)$, in terms of the number of independent sets of its subgraphs.

7 Future Work

In future work, we would like to explore how the concept of periodicity translates from degenerate strings to different families of languages. In particular, we would like to generalize our definitions to apply to *any language*, i.e., any set of strings. We want to investigate which combinatorial results carry over to this more general setting, and if not, which additional conditions must be met.

Moreover, we are interested in studying the algorithmic aspects of the periodicity of languages. One question would be the complexity of determining period sets of degenerate strings; while naïve algorithms are already close to optimal, as shown by the lower bound proven in [14], there might be room for improvement in certain cases, such as the restriction of the alphabet size. A second area of interest is the application of periodicity to matching algorithms on degenerate strings. Similarly to how periodicity is applied to the Knuth-Morris-Pratt algorithm for matching in classical strings, it may be possible to carry over the same concepts to degenerate string matching using our defined terminology for periodicity.

Acknowledgements This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreements No 872539 and No 956229, from the Netherlands Organisation for Scientific Research (NWO) through

Gravitation-grant NETWORKS-024.002.003 and from the Constance van Eeden PhD Fellowship. Moreover, we would like to thank Solon P. Pissis for his helpful advice and suggestions.

References

1. R. Angelo: *A Cameron and Erdős conjecture on counting primitive sets.* INTEGERS, 18 2018, p. 2.
2. F. Blanchet-Sadri: *Algorithmic Combinatorics on Partial Words*, Discrete mathematics and its applications, CRC Press, 2008.
3. F. Blanchet-Sadri, J. Fowler, J. D. Gafni, and K. H. Wilson: *Combinatorics on partial word correlations.* Journal of Combinatorial Theory, Series A, 117(1) 2010, pp. 607–624.
4. F. Blanchet-Sadri, J. D. Gafni, and K. H. Wilson: *Correlations of partial words*, in STACS 2007, 24th Annual Symposium on Theoretical Aspects of Computer Science, Aachen, Germany, February 22-24, 2007, Proceedings, W. Thomas and P. Weil, eds., vol. 4393 of Lecture Notes in Computer Science, Springer, 2007, pp. 97–108.
5. J. C. Bryne, E. Valen, M.-H. E. Tang, T. Marstrand, O. Winther, I. da Piedade, A. Krogh, B. Lenhard, and A. Sandelin: *JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update.* Nucleic Acids Research, 36(suppl 1) 2008, pp. D102–D106.
6. M. Crochemore, C. S. Iliopoulos, T. Kociumaka, J. Radoszewski, W. Rytter, and T. Walen: *Covering problems for partial words and for indeterminate strings.* Theoretical Computer Science, 698 2017, pp. 25–39.
7. P. Erdős: *Note on sequences of integers no one of which is divisible by any other.* Journal of the London Mathematical Society, 10(1) 1935, pp. 126–128.
8. L. Guibas and A. Odlyzko: *Periods in strings.* Journal of Combinatorial Theory, Series A, 30 1981, pp. 19–43.
9. D. Gusfield: *Algorithms on Strings, Trees and Sequences*, Cambridge University Press, 1997.
10. V. Halava, T. Harju, and L. Ilie: *Periods and binary words.* Journal of Combinatorial Theory, Series A, 89(2) 2000, pp. 298–303.
11. J. Holub and W. F. Smyth: *Algorithms on indeterminate strings.* In Proceedings of 14th Australasian Workshop on Combinatorial Algorithms, 2003, pp. 36–45.
12. J. Holub, W. F. Smyth, and S. Wang: *Fast pattern-matching on indeterminate strings.* Journal of Discrete Algorithms, 6(1) 2008, pp. 37–50.
13. C. S. Iliopoulos, R. Kundu, and S. P. Pissis: *Efficient pattern matching in elastic-degenerate strings.* CoRR, abs/1610.08111 2016.
14. C. S. Iliopoulos and J. Radoszewski: *Truly subquadratic-time extension queries and periodicity detection in strings with uncertainties*, in 27th Annual Symposium on Combinatorial Pattern Matching, CPM 2016, June 27-29, 2016, Tel Aviv, Israel, R. Grossi and M. Lewenstein, eds., vol. 54 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016, pp. 8:1–8:12.
15. IUPAC-IUB Commission on Biochemical Nomenclature: *Abbreviations and symbols for the description of the conformation of polypeptide chains. Tentative rules (1969).* Biochemistry, 9(18) 1970, pp. 3471–3479.
16. I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I. Rumynskiy, Y. A. Medvedeva, A. Magana-Mora, V. B. Bajic, D. A. Papatsenko, and et al.: *HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis.* Nucleic Acids Research, 46(D1) Nov 2018, p. D252–D259.
17. H. Liu, P. P. Pach, and R. Palincza: *The number of maximum primitive sets of integers.* Combinatorics, Probability and Computing, 30(5) 2021, p. 781–795.
18. M. Lothaire, ed., *Combinatorics on Words*, Cambridge University Press, second ed., 1997.
19. M. Lothaire: *Algebraic Combinatorics on Words*, Cambridge University Press, Cambridge, 2002.
20. N. McNew: *Counting primitive subsets and other statistics of the divisor graph of 1,2,...,n.* European Journal of Combinatorics, 92 2021, p. 103237.

21. F. Pfeiffer, C. Gröber, M. Blank, K. Händler, M. Beyer, J. L. Schultze, and G. Mayer: *Systematic evaluation of error rates and causes in short samples in next-generation sequencing.* Scientific Reports, 8(1) Jul 2018.
22. N. Pisanti, H. Soldano, and M. Carpentier: *Incremental inference of relational motifs with a degenerate alphabet*, in Combinatorial Pattern Matching, 16th Annual Symposium, CPM 2005, Jeju Island, Korea, June 19-22, 2005, Proceedings, A. Apostolico, M. Crochemore, and K. Park, eds., vol. 3537 of Lecture Notes in Computer Science, Springer, 2005, pp. 229–240.
23. S. Rahmann and E. Rivals: *On the distribution of the number of missing words in random texts.* Combinatorics, Probability and Computing, 12(01) Jan 2003.
24. E. Rivals and S. Rahmann: *Combinatorics of periods in strings.* Journal of Combinatorial Theory, Series A, 104(1) Oct 2003, pp. 95–113.
25. E. Rivals, M. Sweering, and P. Wang: *Convergence of the Number of Period Sets in Strings*, in 50th International Colloquium on Automata, Languages, and Programming (ICALP 2023), K. Etessami, U. Feige, and G. Puppis, eds., vol. 261 of Leibniz International Proceedings in Informatics (LIPIcs), Dagstuhl, Germany, 2023, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, pp. 100:1–100:14.
26. W. F. Smyth and S. Wang: *New perspectives on the prefix array*, in String Processing and Information Retrieval, 15th International Symposium, SPIRE 2008, Melbourne, Australia, November 10-12, 2008. Proceedings, A. Amir, A. Turpin, and A. Moffat, eds., vol. 5280 of Lecture Notes in Computer Science, Springer, 2008, pp. 133–143.
27. H. Soldano, A. Viari, and M. Champesme: *Searching for flexible repeated patterns using a non-transitive similarity relation.* Pattern Recognition Letters, 16(3) 1995, pp. 233–246.
28. The Computational Pan-Genomics Consortium: *Computational pan-genomics: status, promises and challenges.* Briefings in Bioinformatics, 19(1) 2018, pp. 118–135.

A More about lattices

In this appendix, we prove that Ω_n^w, Ω_n^m and Ω_n^s lattices under set intersection and set union, and hence satisfy the Jordan-Dedekind condition. Before we start, we first review some important concepts. We start by recalling the definition of meet and join in terms of posets (partially ordered sets).

Definition (Meet and join). *Given a poset $(A, \leq)$ and $x, y \in A$. We say m is the meet (greatest lower bound or infimum) of x and y denoted by $x \wedge y$, if m satisfies the following conditions.*
1. $m \in A$
2. $m \leq x$ *and* $m \leq y$
3. *For all* $w \in A$, *if* $w \leq x$ *and* $w \leq y$, *then* $w \leq m$.

We say j is the join (least upper bound or supremum) of x and y denoted by $x \vee y$, if j satisfies the following conditions.
1. $j \in A$
2. $x \leq j$ *and* $y \leq j$
3. *For all* $w \in A$, *if* $x \leq w$ *and* $y \leq w$, *then* $j \leq w$.

Definition (Lattice). *Poset $(A, \leq)$ is a lattice if and only if all $x, y \in A$ have both a meet and join.*

Let Ω_n^w, Ω_n^m and Ω_n^s denote the families of weak, medium and strong period sets. In this section, we show that Ω_n^w, Ω_n^m and Ω_n^s are all lattices partially ordered by inclusion.

Theorem 16. *$(\Omega_n^w, \subseteq)$, $(\Omega_n^m, \subseteq)$ and $(\Omega_n^s, \subseteq)$ are lattices with respect to the inclusion order.*

Proof. To show that these posets are lattices, we need to show that their meet and join operations are well-defined. Specifically, since we order their elements with respect to inclusion, we need to show that Ω_n^w, Ω_n^m and Ω_n^s are closed under intersection and union (conditions 2 and 3 are trivially met).

Weak Let $U, V \in \Omega_n^w$ be two weak period sets. Then $0 \in U \subseteq \{0, \ldots, n-1\}$ and $0 \in V \subseteq \{0, \ldots, n-1\}$. It follows that $0 \in U \cup V \subseteq \{0, \ldots, n-1\}$ and $0 \in U \cap V \subseteq \{0, \ldots, n-1\}$. Thus $U \cup V \in \Omega_n^w$ and $U \cap V \in \Omega_n^w$. We conclude that $(\Omega_n^w, \subseteq)$ is a lattice.

Medium Let $U, V \in \Omega_n^m$ be two medium period sets. Equivalently, U and V are two subsets of $\{0, \ldots, n-1\}$ containing 0 and closed under multiplication. It follows that $U \cap V$ and $U \cup V$ also contain 0 and are closed under multiplication. Thus $U \cup V \in \Omega_n^m$ and $U \cap V \in \Omega_n^m$. We conclude that $(\Omega_n^m, \subseteq)$ is a lattice.

Strong Since $\Omega^s = \Omega^m$, the poset of strong period sets $(\Omega_n^s, \subseteq)$ also form a lattice under ordering by inclusion.

We conclude that $(\Omega_n^w, \subseteq)$, $(\Omega_n^m, \subseteq)$ and $(\Omega_n^s, \subseteq)$ are all lattices with respect to the inclusion order. □