Testing Quasiperiodicity

SPIRE 2025

Ben Bals
Centrum Wiskunde & Informatica and VU Amsterdam

Quasiperiodicity

Period	abbaabbaabba
Quasiperiod	abbabbaabba
Seed	abbabbaabba

Strict

Loose

Property Testing

Classic algo

Does a string S have a small quasiperiod?

$\mathcal{O}(n)$ Apostolico, Farach, Iliopoulos, 1991

Property testing

If I can only look at a few positions of S, can I test if it is likely quasiperiodic or not?

If S has small quasiperiod:

Always output YES

If S is far from having a small quasiperiod:

Output NO with $p = 3/4$

need to change at least an ϵ fraction of all letters

Later in the talk

Streaming algorithm

Main Result 1

Theorem. There is a tester deciding whether a string $S \in \Sigma^n$ has a quasiperiod of length at most q using $\mathcal{O}(q^3 \epsilon^{-1} \log q)$ queries, where $\epsilon \in \mathbb{R}^+$ is the distance parameter.

Based on cool combinatorics!

How can a cover self-overlap?

`abbaabba` `abababab`

`abbabba` `ababab`

Write PS(C) for the period lengths of C.

Core insight. If C covers S, then gcd(PS(C)) divides the difference between the occurrences of C in S.

What about the other direction?

~~The Chicken McNugget Numbers~~ Frobenius Numbers

What's the biggest number that's not a sum of 6, 9 and 20?

Chicken McNugget Number

43!

Brady Haran, Numberphile, 2012

Theorem [Erdős & Graham, '72]. Let $A \subseteq \mathbb{N}^+$ be bounded by $q \in \mathbb{N}$. Then any number $x \geq 2q^3$ such that $\gcd(A) | x$ can be written as a conical combination of A.

Tester

Idea 1: Split string into overlapping fragments of length $4q^3$

Observe: If C covers all fragments, then it covers the string

Idea 2: Sample $\tilde{\mathcal{O}}(q/\epsilon)$ fragments and check that C

(1) covers of all of them, and

(2) the indices align.

Easy: If C is a cover, this always accepts.

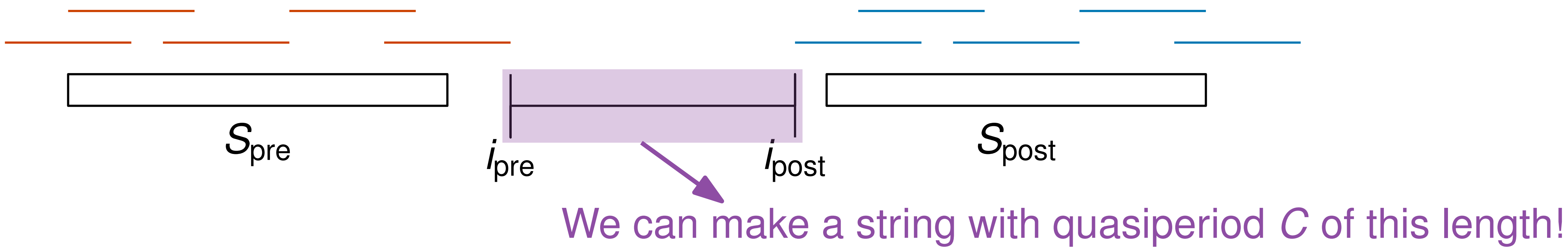
Harder: Why does it reject if S is ϵ-far from having a small quasiperiod?

Lemma. If S is ϵ-far from $\mathrm{QP}_\Sigma(q)$, for some $q \in \mathbb{N}$, the set $\mathcal{S}' \subseteq \mathcal{S}$ of C-consistent fragments of S has size at most $\mathcal{O}((1-\epsilon/2) \cdot n/q^3)$.

Applying the Nuggets

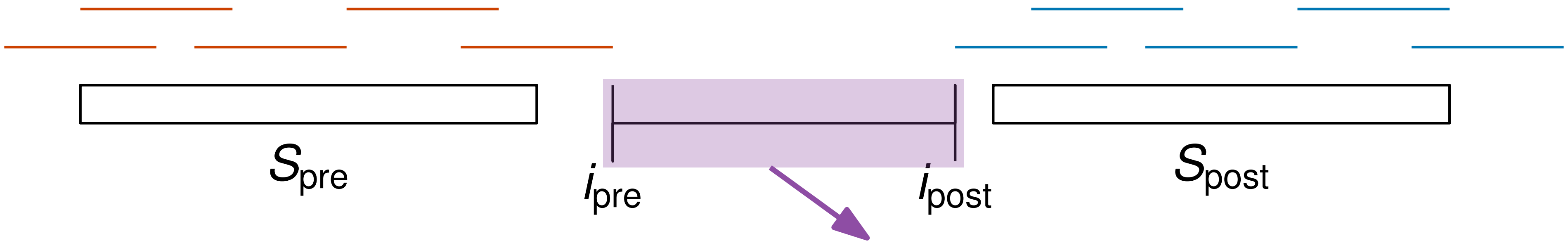
Lemma. If S is ϵ-far from $\mathrm{QP}_\Sigma(q)$, for some $q \in \mathbb{N}$, the set $\mathcal{S}' \subseteq \mathcal{S}$ of C-consistent fragments of S has size at most $\mathcal{O}((1-\epsilon/2)\cdot n/q^3)$.

- For contradiction: Assume there are few C-inconsistent fragments $\mathcal{T}$
  - $|\mathcal{T}| \leq \frac{\epsilon n}{4q^3}$
- Goal: Then S is not ϵ-far from having a small quasiperiod
- Strategy: Iterate over all inconsistent fragments and fix them



Applying the Nuggets (2)

Lemma. If S is ϵ-far from $\mathrm{QP}_\Sigma(q)$, for some $q \in \mathbb{N}$, the set $\mathcal{S}' \subseteq \mathcal{S}$ of C-consistent fragments of S has size at most $\mathcal{O}((1-\epsilon/2) \cdot n/q^3)$.



We can make a string with quasiperiod C of this length!

Theorem [Erdős & Graham, '72]. Let $A \subseteq \mathbb{N}^+$ be bounded by $q \in \mathbb{N}$. Then any number $x \geq 2q^3$ such that $\gcd(A) | x$ can be written as a conical combination of A.

By choice of the fragments — By C-consistency

Theorem.
Let $x \geq 2q^3$ and $\gcd(\mathrm{PS}(C)) | x$, then there is a string of length x with quasiperiod C.

■-ish

Streaming!

Theorem. For any string $S \in \Sigma^n$ and any $q \in \mathbb{N}$, there is a one-pass streaming algorithm for computing the shortest cover C of S, if $|C| \leq q$, that uses $\mathcal{O}(q)$ space and runs in $\mathcal{O}(n)$ time w.h.p.

Randomness comes from an internal data structure, we don't sample here!

State of the art: $O(\sqrt{n \log n})$ space and $O(n \log^2 n)$ time

Gawrychowski, Radoszewski, Starikovskaya, 2019

Take aways

Combinatorial Insights for Better Algorithms

[Erdős & Graham, ‘72]

Have Fun

Insights Often Work In Many Models

Property Testing → Streaming

References

Brady Haran. How to order 43 Chicken McNuggets (Frobenius numbers) - Numberphile, 2012. https://www.youtube.com/watch?v=vNTSugyS038.

Paul Erdős and Ronald Graham. On a linear Diophantine problem of Frobenius. Acta Arithmetica, 21:399–408, 1972.

Pawel Gawrychowski, Jakub Radoszewski, and Tatiana Starikovskaya. Quasi-periodicity in streams. CPM, 2019.

Oded Lachish and Ilan Newman. Testing periodicity. Algorithmica, 60(2):401–420, 2011.

Alberto Apostolico, Martin Farach, and Costas S. Iliopoulos. Optimal superprimitivity testing for strings. Inf. Process. Lett., 39(1):17–20, 1991.

Bonus: Overlaps