CS5800 – ALGORITHMS

MODULE 1. REVIEW OF ASYMPTOTIC NOTATION

Lesson 3: PCR

Ravi Sundaram

Topics

- What is PCR?
- Why PCR?
- Basics of biochemistry
- An algorithmic puzzle
- Summary

What is PCR?

- Polymerase Chain Reaction
  - technique to generate millions of copies of a particular DNA sequence from a single piece of DNA
  - Chain Reaction – nuclear 1 to 2 to 4 to 8 to explosion
  - Polymerase enzyme that enables chain reaction
- Kary Mullis won the 1993 Biology Nobel for discovering PCR
- Revolutionary – Before PCR and After PCR
  - Before: DNA cloning – back-breaking, countless lab hours, low yield
  - After: CSI, 23andMe

Why PCR?

- Timely
  - In the news thanks to Covid
- At its heart it is an algorithm
  - Should have been discovered by a computer scientist
- Polynomial vs Exponential
  - Exemplary demonstration of exponential growth rate

Basics of biochemistry

- Deoxyribonucleic acid (DNA)
  - a nucleic acid containing genetic instructions used in development/ functioning of most known living organisms.
  - Segments carrying genetic information are called genes.
  - Consists of two long polymers of simple units called nucleotides. These two strands run in opposite directions to each other.
  - Each nucleotide contains one of four types of molecules called bases (or nucleobases or nucleic acids). These four bases are denoted: A (adenine), T (thymine), C (cytosine) and G (guanine).

A→T→T→C→A→G→T

Basics of biochemistry (continued)

- Within cells DNA is organized into long structures called chromosomes
- Characterized by the now-legendary Watson-Crick double-helix structure comprised of two anti-parallel strands with the weak bonds: A –T and C–G.

A→T→T→C→A→G→T

| | | | | | |

T←A←A←G←T←C←A

Puzzle: 4 well-known facts

- **Heating**
  - when a test tube of double stranded DNA is **heated** (in a controlled fashion) then it breaks up into two constituent and complementary single strands
- **Cooling**
  - when single strands of DNA are **cooled** then complementary copies will rejoin
- **Polymerase**
  - when a strand of DNA is cooled in a test tube with an abundant supply of bases along with **polymerase** then polymerase will create a complementary strand by moving along the given strand and adding bases one by one.

A→T→T→C→A→G→T

| |

C←A

Puzzle: 4 well-known facts

- Polymerase will only add in the complementary direction to the given strand:

A→T→T→C→A→G→T

| | |

T←A←A

- Complementary strands started by polymerase are not stable until they are many bases long. Addition of polymerase will only rarely result in a complementary strand that starts at the beginning and goes all the way to the end.

- **Primers -** short DNA strands (tens of bases) obtainable in quantities through conventional (pre-PCR) techniques.

Algorithmic Puzzle

- given a small quantity of DNA containing a DNA subsequence of interest, create many copies of the DNA subsequence.

PCR - Kary Mullis' algorithm

Let the given double strand be

……..→o→……..→ω→…….

| |

……..←ó←……..←ώ←…….

and let the sequence of interest be bracketed between o and ω, i.e. o→……..→ω.

add enough primers o and ώ and then to heat and cool repeatedly

2-sided and 1-sided strands

- after the first heating:

……..→o→……..→ω→…….

and

……..←ó←……..←ώ←…….

- Now comes the magic of polymerase with the primers, and we get:

……..→o→……..→ω→…….
| |
……..←ó←……..←ώ

And

o→……..→ω→…….
| |
……..←ó←……..←ώ←…….

Terminal strands appear

After the second heating:

……..→o→……..→ω→…….

and

……..←ó←……..←ώ

and

o→……..→ω→…….

and

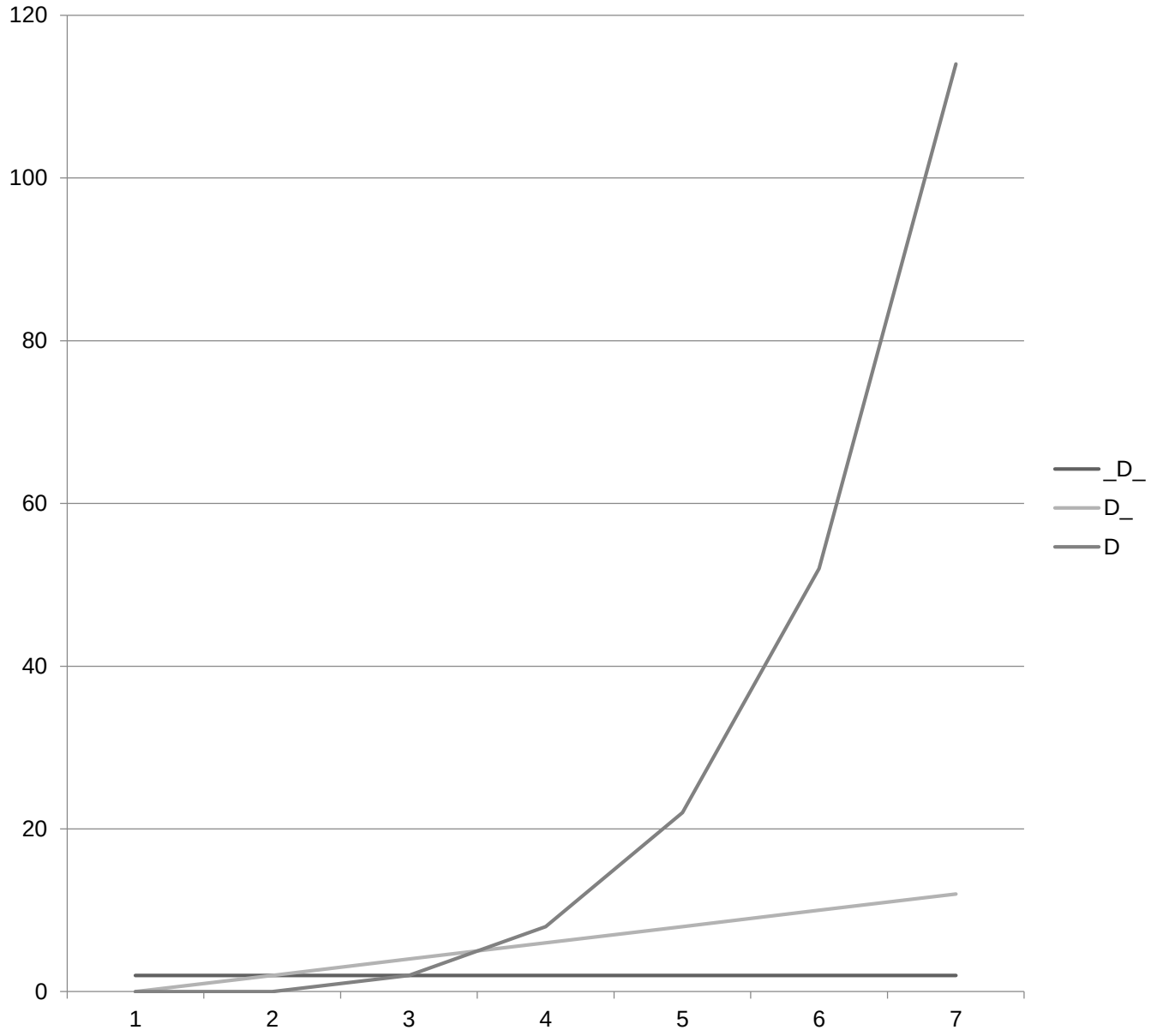
……..←ó←……..←ώ←…….

Distribution of strands – 2-sided, 1-sided and terminal

- Let D denote desired seq or complement, and _ denote extension

Heating cycle	_D_	D_	D
1	2	0	0
2	2	2	0
3	2	4	2
4	2	6	8
5	2	8	22
n	2	2(n-1)	2^n-2n

Distribution of strands – exponential vs polynomial(linear) vs constant



Summary

- Algorithms are central to the functioning of modern-day society
- The time complexity of an algorithm is a worst-case measure, parameterized in terms of input size and expressed in order notation which is asymptotic and ignores constants
- Main Takeaway: Exponential growth dramatically outpaces polynomial growth, hence the emphasis on discovering and using polynomial-time algorithms