Course: Algorithms and Applications (15-499)

Instructor: Guy Blelloch

TA: Shuchi Chawla

URL:
www.cs.cmu.edu/~guyb/realworld/indexS03.html

Expected Background

Asymptotic analysis (solving recurrences, NP-complete problems)

Graph Terminology and Algorithms (e.g. min-spanning-tree and max-flow algorithms)

Dynamic programming

Matrix algebra (e.g. null-spaces, eigenvalues)

Fast Fourier Transform

Probability

Course Requirements

Readings

- Chapters, papers, course-notes, web documents
- There is NO course textbook

Assignments

- 6 written assignments (must hand them in individually)

Project

- Group programming project due at the end of the semester. Groups of 2 or 3.

Midterm and Final

Why care about algorithms?

Just about every field uses algorithms

- Engineering, Physical sciences, Social sciences
- Business, Medicine
- Probably used more widely than Calculus

The level of sophistication has increased significantly in recent years

Asymptotics becoming more important

- Linear programming (interior-point methods)
- Factoring (number-field sieve)
- …

Some Examples

Standard complaints about algorithm analysis

Irrelevant problem

Impractical constants hidden in big-O

Bad machine models

Too complicated

On Strassen's Matrix Multiply

From Numerical Recipes (Press et. al. 1986)

> "We close the chapter with a little entertainment, a bit of algorithmic pretidigitation

and later in the same section

> "This is fun, but let's look at practicabilities: If you estimate how large N has to be before the difference between exponent 3 and exponent 2.807 is substatial enough to outweigh the bookkeeping overhead, arising from eht compicated nature of the recursive Strassen algorithm, you will find that LU decomposition is in no immediate danger of coming obsolete."

From an Experimental Paper

From an experimental paper (Bailey, Lee and Simon):

> "For many years it was thought that [the level at which Strassen's algorithm is more efficient] was well over 1000x1000. In fact, for some new workstations, such as the Sun-4 and SGI IRIS 4D, Strassen is faster for matrices as small as 16x16.

Most modern matrix libraries now supply Strassen as a choice

- IBM ESSL library
- Cray Libraries
- LAPACK library

Topics

Data Compression (5)

- gzip, bzip, jpeg, mpeg, …

Cryptography (5)

- ssh, RSA, Rijdael, kerberos, …

Error Correcting Codes (3)

- Reed-Solomon, tornado-codes, CDs

Computational Biology (5)

- Genome sequencing, BLAST, evolutionary trees, …

Indexing and Searching (5)

- Google pagerank, duplicate removal, …

(n) = number of lectures (approximate)

What each topic will cover

Key problems in the area

- Formal definitions
- How they relate to practice

Many algorithms

- Theory
- Practice

Key applications along with case studies

Some General Themes

Measures of "performance" beyond "Time"

- Quality of results
- Generality
- Simplicity

Transition from theory to practice

- Often an iterative process

Characterizing input

- Input data is rarely "worst case" or even "expected case" when the expectation is over all possible inputs.

For Next Tuesday

Read the readings on Data Compression:

http://www.cs.cmu.edu/~guyb/realworld/indexS03.html

Start homework 1a.