CST1
COMPUTER SCIENCE TRIPOS Part IB

Wednesday 7 June 2023 13:30 to 16:30

COMPUTER SCIENCE Paper 6

*Answer **five** questions.*

*Submit the answers in five **separate** bundles, each with its own cover sheet. On each cover sheet, write the numbers of **all** attempted questions, and circle the number of the question attached.*

You may not start to read the questions printed on the subsequent pages of this question paper until instructed that you may do so by the Invigilator

STATIONERY REQUIREMENTS
Script paper
Blue cover sheets
Tags

SPECIAL REQUIREMENTS
Approved calculator permitted

1 Complexity Theory

Recall that a (*simple, undirected*) graph G is a set of vertices V along with a set of edges E, where each edge $e \in E$ is a two-element subset of V. For the purpose of this question, all graphs are simple, undirected graphs.

For a graph $G = (V, E)$ and an edge $e \in E$, we write $G - e$ to denote the graph obtained from G by *removing* the edge e. That is $G - e$ has exactly the same vertices as G and all edges in E except for e.

(a) What is a *Hamiltonian cycle* in a graph $G = (V, E)$? [2 marks]

(b) What is known about the complexity of deciding whether a given graph G has a Hamiltonian cycle? [2 marks]

(c) Show that G has a Hamiltonian cycle that does not include the edge e if, and only if, $G - e$ has a Hamiltonian cycle. [4 marks]

(d) Assume that P=NP. Using this assumption, show that there is a polynomial-time algorithm A such that if A is given a graph $G = (V, E)$, it will return "no" if G does not contain a Hamiltonian cycle and return a Hamiltonian cycle of G otherwise. [12 marks]

2 Complexity Theory

Let $f : \Sigma^* \to \Sigma^*$ be a function on Σ-strings for some finite alphabet Σ. Say that f is a *pseudo one-way function* if it satisfies the following three conditions:

- There is a constant k such that for every $x \in \Sigma^+$, $|x|^{1/k} \leq |f(x)| \leq |x|^k$. (Here $|x|$ denotes the length of a string x).
- f is computable by a polynomial-time algorithm.
- There is no function g, computable in polynomial time, such that $f(g(y)) = y$ for all strings y in the range (i.e. image) of f.

For a pseudo one-way function f, let $L_f \subseteq \Sigma^* \times \Sigma^*$ be the following set

$$L_f = \{(x, y) \mid \exists z(z \leq_{\text{lex}} x \text{ and } f(z) = y)\}.$$

Here $\leq_{\text{lex}}$ denotes the lexicographic order on strings.

(*a*) How would you modify the definition of a pseudo one-way function to obtain the definition of a *one-way function* in the sense defined by Papadimitriou? [3 marks]

(*b*) Show that for any pseudo one-way function f, the language L_f is in NP. [4 marks]

(*c*) Show that for any pseudo one-way function f, the language L_f is not in P. [4 marks]

In the following, ϕ denotes an arbitrary Boolean formula and T a list assigning a Boolean value to each variable appearing in ϕ. Fix Σ to be a suitable alphabet in which we can write ϕ and T as well as the string "no" and consider the following function defined on all Σ-strings.

$$s(x) = \begin{cases} \phi & \text{if } x = (\phi, T) \text{ and } T \text{ satisfies } \phi \\ \text{“no”}x & \text{otherwise} \end{cases}$$

(*d*) Prove that if P≠NP, then s is a pseudo one-way function. [9 marks]

3 Computation Theory

(*a*) Define what is meant by a *configuration* and a *computation* of a register machine (RM). Explain carefully what it means to say that a computation *halts*. [5 marks]

(*b*) Define the notion of RM *computable* partial numerical function of n arguments. [2 marks]

(*c*) What does it mean for a problem (expressed as a property of numbers) to be RM *undecidable*? [2 marks]

(*d*) A computation of a RM is said to be *circular* if it reaches the same configuration at two different times.

(*i*) Explain why a circular computation does not halt. Give an example of a RM computation that does not halt, but that is not circular. [2 marks]

(*ii*) The Circularity Problem is: Decide whether or not the computation of any given RM and initial register contents is circular. Give a proof that the Circularity Problem is undecidable. You may assume suitable functions for encoding and decoding pairs of numbers as numbers, finite lists of numbers as numbers, and RM programs as numbers. [9 marks]

4 Computation Theory

(*a*) For the λ-calculus, define the notions of

(*i*) *β-conversion* ($=_\beta$) [2 marks]

(*ii*) *Church numeral* ($\underline{n}$) [2 marks]

(*b*) What does it mean for a total function $f : \mathbb{N}^n \to \mathbb{N}$ to be *λ-definable*? Explain why it is the case that not every $f : \mathbb{N}^n \to \mathbb{N}$ is λ-definable, carefully stating any standard results that you rely upon. [3 marks]

(*c*) Explain why the predecessor function $pred : \mathbb{N} \to \mathbb{N}$

$$pred(x) = \begin{cases} 0 & \text{if } x = 0 \\ x - 1 & \text{if } x > 0 \end{cases}$$

is λ-definable and give a λ-term that represents it. [4 marks]

(*d*) Show that the following functions are λ-definable. For each part you may assume solutions to the previous parts of the question.

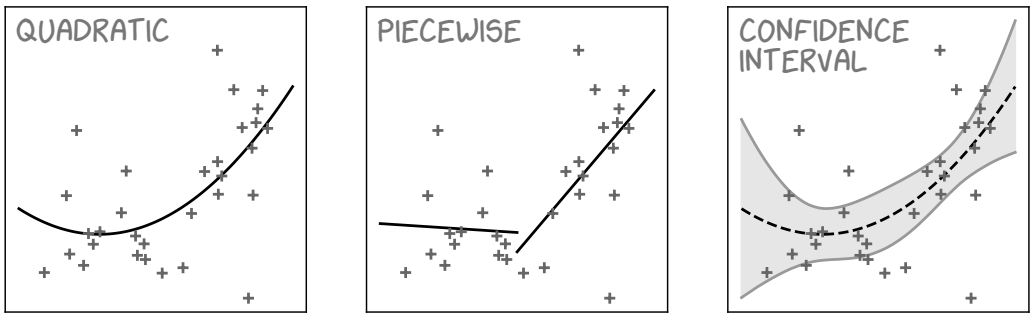
(*i*) $if_0 : \mathbb{N}^3 \to \mathbb{N}$, where $if_0(x, y, z) = \begin{cases} y & \text{if } x = 0 \\ z & \text{if } x \neq 0 \end{cases}$ [3 marks]

(*ii*) $and : \mathbb{N}^2 \to \mathbb{N}$, where $and(x, y) = \begin{cases} 0 & \text{if } x = 0 \text{ and } y = 0 \\ 1 & \text{if } x \neq 0 \text{ or } y \neq 0 \end{cases}$ [1 mark]

(*iii*) $monus : \mathbb{N}^2 \to \mathbb{N}$, where $monus(x, y) = \begin{cases} x - y & \text{if } x > y \\ 0 & \text{if } x \leq y \end{cases}$ [3 marks]

(*iv*) $eq : \mathbb{N}^2 \to \mathbb{N}$, where $eq(x, y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$ [2 marks]

5 Data Science



These three plots are adapted from xkcd comic 2048 "Curve-fitting methods and the messages they send" by Randall Munroe (CC BY-NC 2.5)[†].

(*a*) Describe a probability model and a fitting procedure that might be behind the "Quadratic" plot. Give pseudocode for the fitting procedure. Explain briefly how the plot is produced. [*Note:* When describing your model, remember to state any relevant probability distributions.] [7 marks]

(*b*) Repeat part (a) for the "Piecewise" plot. [5 marks]

(*c*) Repeat part (a) for the "Confidence interval" plot. In your answer, explain carefully the basis for your confidence interval calculation. [8 marks]

[†] `https://creativecommons.org/licenses/by-nc/2.5/legalcode`

6 Data Science

Here are two binary sequences:

$$x = \texttt{111010100100001011110011}$$
$$y = \texttt{000000011111111111110000}$$

(*a*) Consider a model in which each element of x is an independent Bernoulli random variable, $\mathrm{Bin}(1, p)$. Estimate p, and give a formula for the log likelihood of the x sequence. Repeat for y. Explain why the log likelihoods are the same for the two sequences. [4 marks]

(*b*) The model from part (*a*) might seem a poor choice for y. Explain how to conduct a hypothesis test to determine whether this is so. In your answer you should define the test statistic you have invented for the test, and you should explain your choice of one-sided or two-sided testing. [8 marks]

(*c*) Consider an alternative model, in which y is generated from a two-state Markov chain, with the first item of y drawn from the chain's stationary distribution. Give a formula for the log likelihood of y, and explain how to fit the model. [8 marks]

[*Note:* You do not need to find numerical answers. You may be interested to know that for the independent model the log likelihoods are both -16.55, and for the Markov model the log likelihood of x is -16.47 and that for y is -9.00.]

7 Logic and Proof

(*a*) Exhibit a model for the following set of formulas, or prove that none exists. Briefly explain your work in each step.

$$P \qquad P \to (R \to Q) \qquad P \vee \neg Q \vee \neg P \qquad Q \to S \wedge \neg T \qquad S \to Q \vee T$$

[6 marks]

(*b*) For each of the following sets of formulas, either exhibit an interpretation in S4 modal logic that satisfies them simultaneously at a particular world, w, or show through a formal proof that they cannot be satisfied.

(*i*) $\Diamond\Box P, \quad Q, \quad \Box\Diamond\Box\neg Q, \quad \Box(P \to \Diamond R \wedge \Diamond\neg R), \quad \Box(\Box\neg Q \vee \neg\Diamond P)$

[8 marks]

(*ii*) $\Box(P \vee Q), \quad \Diamond\neg P, \quad \neg\Diamond Q$

[6 marks]

8 Logic and Proof

Which of the following formulas is valid? Prove them using resolution or find a counterexample. If a counterexample exists, explain how resolution is blocked from producing a spurious proof.

(*a*) $(\forall x\,[\neg Q(x,x) \to P(x)]) \to \forall x\,[P(x) \vee \exists y\; Q(x,y)]$ [7 marks]

(*b*) $[\exists x\, P(x)] \to \exists x\, P(f(x))$ [6 marks]

(*c*) $\exists z\, \forall w\, [(\forall x\,[P(x) \to \exists y\, Q(x,y)]) \to (P(w) \to Q(w,z))]$ [7 marks]

9 Semantics of Programming Languages

The relational algebra is a small language for manipulating sets of tuples, and is one of the central objects of study in database theory. We can give a syntax for (a subset of) it as follows:

τ	::=	$\mathsf{int} \mid \mathsf{bool}$	Data types
d	::=	$n \mid b$	Data values
R	::=	$[l_1 : \tau_1, \ldots, l_n : \tau_n]$	Record types (with disjoint field names l_i)
r	::=	$[l_1 = v_1, \ldots, l_n = v_n]$	Record values (with disjoint field names l_i)
S	::=	$\mathsf{Set}\ R$	Set types
e	::=	$\{r_1, \ldots, r_n\}$	Set literal
	$\mid$	$e \cup e'$	Set union
	$\mid$	$e \times e'$	Cartesian product with disjoint field labels
	$\mid$	$\Pi_{l_1,\ldots,l_n}(e)$	Records of e with fields not in $l_1, \ldots, l_n$ removed
	$\mid$	$\sigma_{l_1=l_2}(e)$	Subset of e where the fields l_1 and l_2 are equal

(*a*) State the form of the typing judgements for this language, and give typing rules for this programming language ascribing to each category of terms its corresponding types. [8 marks]

(*b*) Define a deterministic small-step operational semantics for this language, defining any auxiliary functions you need as well. [10 marks]

(*c*) Give a precise statement of the progress and preservation properties for this language. You do not need to give a proof. [2 marks]

10 Semantics of Programming Languages

Consider the language with functions, integers, and printing.

$$\begin{array}{lcll} \tau & ::= & \mathsf{unit} \mid \mathsf{int} \mid \tau \to \tau' & \text{Types} \\ e & ::= & x \mid \lambda x : \tau.\, e \mid e\, e' \mid \mathsf{skip} \mid n \mid \mathsf{print}(e) \mid e; e' & \text{Terms} \end{array}$$

The typing rule for $\mathsf{print}(e)$ is:

$$\frac{\Gamma \vdash e : \mathsf{int}}{\Gamma \vdash \mathsf{print}(e) : \mathsf{unit}}$$

(*a*) Define a small-step, call-by-value operational semantics for this language. Clearly explain what the components of the machine configuration are, and how it identifies what is printed. [10 marks]

(*b*) State a progress theorem for this language, and explain what it says about the evolution of the machine state. [4 marks]

(*c*) Prove progress for the $\mathsf{print}(e)$ case, giving the names of any of the standard properties (such as substitution) that you needed to use in the proof. [6 marks]

END OF PAPER