UNIVERSITY OF CAMBRIDGE

COMPUTER SCIENCE TRIPOS Part II

Tuesday 31 May 2016 1.30 to 4.30

COMPUTER SCIENCE Paper 7

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

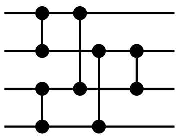
SPECIAL REQUIREMENTS
Approved calculator permitted

This page is intentionally left blank.

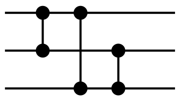
1 Advanced Algorithms

(*a*) State the zero-one principle in the context of sorting networks. [2 marks]

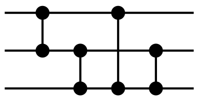
(*b*) For each of the following six comparison networks, state whether it is a sorting network or not. In each case, justify your answer. For the justification you may refer to standard results without giving a proof. [9 marks]



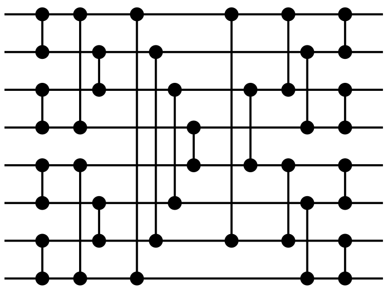
(1)



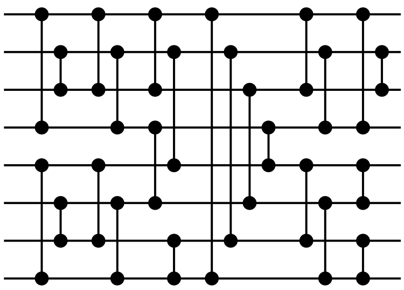
(2)



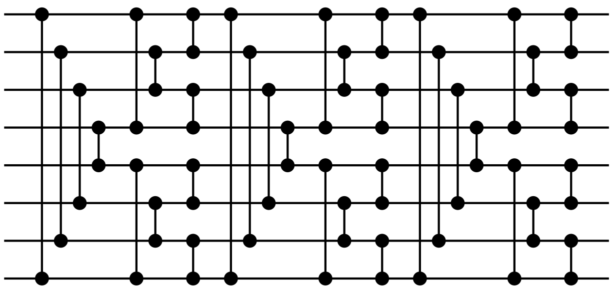
(3)



(4)



(5)



(6)

(*c*) Let n be an exact power of 2. Show how to construct an n-input, n-output comparison network of depth $\log n$ in which the top output wire always carries the minimum input value and the bottom output wire always carries the maximum input value. [4 marks]

(*d*) (*i*) Prove that the number of comparators in any sorting network is $\Omega(n \log n)$. [4 marks]

(*ii*) What does Part (*d*)(*i*) imply in terms of the depth of any sorting network? [1 mark]

2 Advanced Graphics

A *force function* $F : \mathbb{R}^3 \rightarrow \mathbb{R}$ takes a 3D point and returns a scalar representing a value of force. Force functions are the fundamental building blocks of metaball modelling.

We will build an implicit surface renderer which takes as input a set of force functions $\{F_1(P), \ldots, F_n(P)\}$ and renders the set of all points P in space where the forces of the functions sum to a threshold: the 3D isosurface such that $\sum F_i(P) = 0.5$.

(a) Using pseudocode, give a force function *Sphere(P)* which will render a unit sphere centred on $(0, 0, 0)$. [Figure 1] [2 marks]

(b) Using pseudocode, give a force function *Cube(P)* which will render an axis-aligned cube of edge length 2 centred on $(1, 1, -1)$. [Figure 2] [4 marks]

(c) You now pass both *Sphere(P)* and *Cube(P)* to your implicit surface renderer. Depending on your choice of force functions, the seam between the cube and the sphere may be a sharp edge (to within the tolerance of your polygonalization) or a smooth blend which merges gradually from one form into the other. Which will it be, and (briefly) why? [Figures 3 and 4] [2 marks]

(d) Provide alternate formulations of *Sphere(P)* and/or *Cube(P)* such that if you answered 'smooth' to Part (c) then your answer would now be 'sharp', or vice-versa. [4 marks]

A *spatial distortion function* $S : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ transforms one 3D point to another. If the points passed into the force function are modified by a spatial distortion function—that is, if we render $F(S(P))$—then the rendered isosurface will have a different shape.

For example, if we define $S(P)$ as

```
function Point S(P) {
  return new Point(P.x * 2, P.y / 2, P.z * 2);
}
```

then rendering the implicit surface of *Sphere(S(P))* will yield a tall, narrow ellipsoid along the Y axis. [Figure 5]

(e) Give a spatial distortion function $S(P)$ such that rendering the isosurface of *Cube(S(P))* would render the cube centred at the origin and rotated 45 degrees around the X axis. [Figure 6]

Hint: a standard rotation matrix is $\begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix}$. [3 marks]

(*f*) Define $S(P)$ as

```
function Point S(P) {
  return new Point(
      P.x / 4,
      P.y * 2 / sin(P.x * PI),
      P.z * 2);
}
```

Describe and draw a sketch of the isosurface defined by *Sphere(S(P))*.

[5 marks]

Figures:

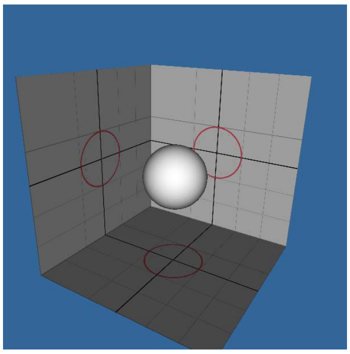


Figure 1: A sphere centred at $(0, 0, 0)$

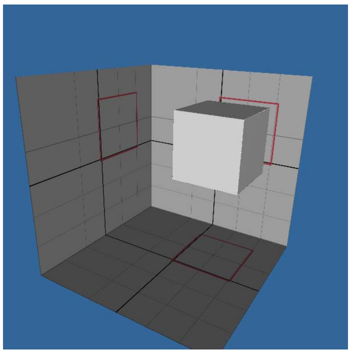


Figure 2: A cube of edge length 2 centred at $(1, 1, -1)$

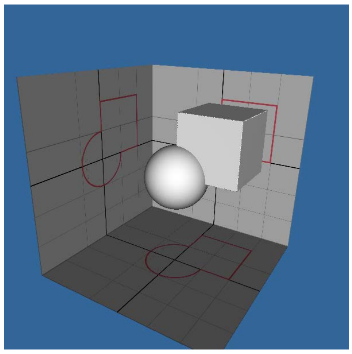


Figure 3: A sharp join between sphere and cube

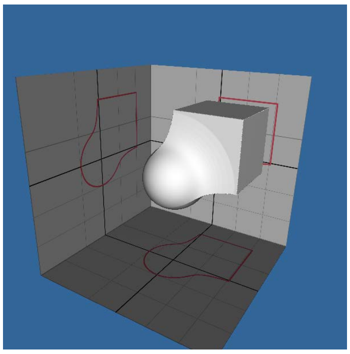


Figure 4: A smooth blending between sphere and cube

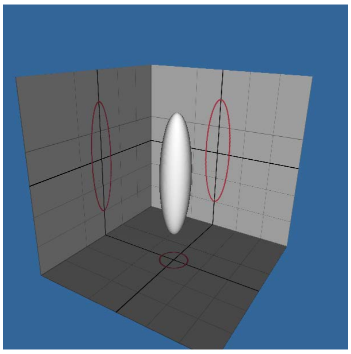


Figure 5: A vertical ellipsoid

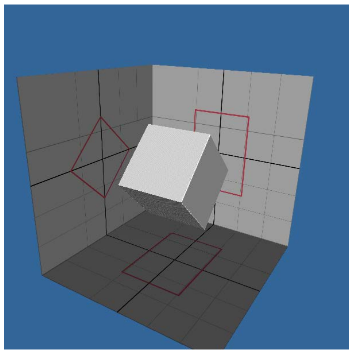


Figure 6: A tilted cube centred at $(0, 0, 0)$

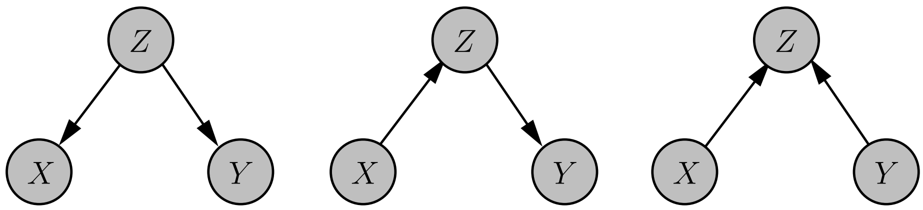
3 Artificial Intelligence II

Let X, Y and Z be random variables. Denote by $X \perp Y \mid Z$ that X and Y are conditionally independent given Z.

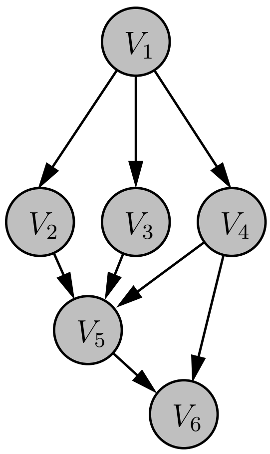
(*a*) Give *two* definitions of what it means to say that $X \perp Y \mid Z$, and prove that they are equivalent. [4 marks]

(*b*) Prove that if $X_1, X_2 \perp Y_1, Y_2 \mid Z$ then $X_1, X_2 \perp Y_1 \mid Z$. [3 marks]

(*c*) For each of the following Bayesian networks, state whether $X \perp Y \mid Z$ and justify your answer. [7 marks]



(*d*) Give a detailed explanation of how a *Markov chain Monte Carlo algorithm* might be used to estimate an arbitrary inference for the following Bayesian network. [6 marks]



4 Bioinformatics

(*a*) Explain the uses of Eulerian and Hamiltonian graphs in the context of genome assembly. [4 marks]

(*b*) Discuss, giving an example, how to apply de Bruijn graphs to genome assembly. [6 marks]

(*c*) Discuss how the choice of different K-mer length affects the accuracy of genome reconstruction. [3 marks]

(*d*) Discuss the additive property in phylogeny. [3 marks]

(*e*) Show one example of additive and one of non-additive matrices. [4 marks]

5 Business Studies

(*a*) A small software company accepts a contract to supply some specialist software. The contract is worth £2M, paid by payments of £500k at start and three payments of £500k invoiced against milestones expected at months 1, 3 and 5 of the 6 month project. Staff costs are estimated at £150k/month and overheads of £100k/month. Draw up an outline cashflow for this project. [5 marks]

(*b*) How profitable is this project if all goes to plan? [2 marks]

(*c*) At a project beer night towards the end of Month 3, your lead programmer punches the project manager in the face, during a heated argument. The project manager comes to you on the following day to complain to you in your role as CEO. How will you handle this affair and what is the likely effect on the project's progress? [8 marks]

(*d*) How would the resulting loss of progress affect the project's profitability? [5 marks]

6 Comparative Architectures

A large Last Level Cache (LLC) is necessary to achieve good performance in many applications. Recent server class processors have included LLCs with capacities of 40 MBytes or more. Large caches such as this are constructed from numerous smaller SRAM banks.

(*a*) Describe an appropriate on-chip network to interconnect 32 SRAM banks to create a large LLC. The delay to access a bank should increase as we move further away from the cache controller and bus interface. The SRAM banks are square and the time taken for a signal to travel along the edge of a SRAM bank is much less than your network's clock cycle time. [5 marks]

(*b*) To implement a set-associative LLC we may spread each set across multiple banks, i.e. each "way" of the set will be in a different bank. The different associative ways will have different access latencies depending on their distance from the cache controller. How might we optimise the placement of lines in particular banks (or ways) to minimise the cache's average access latency? Remember to consider the cost of moving lines. [6 marks]

(*c*) How might the SRAM banks be efficiently interconnected so that the cache's access time is constant regardless of which bank is accessed? [4 marks]

(*d*) Why might it be advantageous to be able to manage the amount of LLC used by each co-scheduled thread in a chip multiprocessor? [5 marks]

7 Denotational Semantics

(*a*) (*i*) Define the notion of continuous function between domains. [2 marks]

(*ii*) Let $\mathcal{P}(\mathbb{N}^2)$ be the domain of all subsets of pairs of natural numbers ordered by inclusion. Show that the function $f : \mathcal{P}(\mathbb{N}^2) \to \mathcal{P}(\mathbb{N}^2)$ given by

$$f(S) = \{\, (1,1) \,\} \cup \{\, (x+1, x \cdot y) \in \mathbb{N}^2 \mid (x,y) \in S \,\} \qquad (S \subseteq \mathbb{N}^2)$$

is continuous. [3 marks]

(*b*) (*i*) State Tarski's fixed point theorem for a continuous endofunction on a domain. [2 marks]

(*ii*) Give a concrete explicit description of the fixed point $\mathit{fix}(f) \subseteq \mathbb{N}^2$ of the continuous function f in Part (*a*)(*ii*). Briefly justify your answer. [3 marks]

(*c*) (*i*) Define the notion of an admissible subset of a domain. [2 marks]

(*ii*) Let $P \subseteq \mathcal{P}(\mathbb{N}^2)$ be defined as $P = \{\, S \subseteq \mathbb{N}^2 \mid \forall\,(x,y) \in S.\ \log y \leq x \cdot \log x \,\}$. Show that P is an admissible subset of the domain $\mathcal{P}(\mathbb{N}^2)$. [3 marks]

(*d*) (*i*) State Scott's fixed point induction principle. [2 marks]

(*ii*) Use Scott's fixed point induction principle to show that $\mathit{fix}(f) \in P$ for f the continuous function in Part (*a*)(*ii*) and P the admissible subset of the domain $\mathcal{P}(\mathbb{N}^2)$ in Part (*c*)(*ii*). [3 marks]

8 Hoare Logic and Model Checking

This question considers a language $\mathcal{L}$ which has integer variables V, arithmetic expressions E and boolean expressions B, along with commands C of the forms V:=E (assignment), $C;C'$ (sequencing), `IF` B `THEN` C `ELSE` C' (conditional) and `WHILE` B `DO` C (iteration).

(*a*) Explain the syntax of the Hoare-logic partial-correctness formula $\{P\}\ C\ \{Q\}$ and give a careful definition in English of when it is valid, that is, when $\models \{P\}\ C\ \{Q\}$. [2 marks]

(*b*) How does the definition of validity for the total-correctness formula $[P]\ C\ [Q]$ differ? [1 mark]

(*c*) Preconditions and postconditions in $\{P\}\ C\ \{Q\}$ often make use of logical or auxiliary variables v in addition to program variables V. Explain why this is useful illustrating your answer with a command C which satisfies $\{\mathbf{T}\}\ C\ \{\mathtt{R} = \mathtt{X} + \mathtt{Y}\}$ but not $\{\mathtt{X} = x \wedge \mathtt{Y} = y\}\ C\ \{\mathtt{R} = x + y\}$. [3 marks]

(*d*) Give the axioms and rules of an inference system $\vdash \{P\}\ C\ \{Q\}$ for Hoare logic. [4 marks]

(*e*) Are your rules sound? To what extent are they complete? [2 marks]

(*f*) Give a formal proof, using your inference system, of $\{\mathtt{X} = x \wedge \mathtt{Y} = 3\}$ `X:=X+1` $\{\mathtt{X} - 1 = x \wedge \mathtt{Y} < 10\}$. [2 marks]

(*g*) Consider the command C given by `WHILE X>0 DO (X:=X-1; Y:=Y+3)`, and let P be the precondition $\mathtt{X} = x \wedge \mathtt{Y} = y \wedge x \geq 0$. Give the strongest postcondition Q that you can establish. Give any invariant necessary to prove $\{P\}\ C\ \{Q\}$ for your Q. Explain briefly how the structure of the proof relates to the structure of C. [6 marks]

9 Human–Computer Interaction

This question is concerned with methods that might contribute to the design of a novel wearable device, which can provide unobtrusive cues to the user, helping them to negotiate social situations. The device is configured by the user with a web interface to an XML database that specifies many different types of social obligations and challenges that the user wishes to deal with. It also uses artificial intelligence techniques to predict appropriate actions based on the user's social media history.

(*a*) Explain the difference between formative and summative HCI research methods. [2 marks]

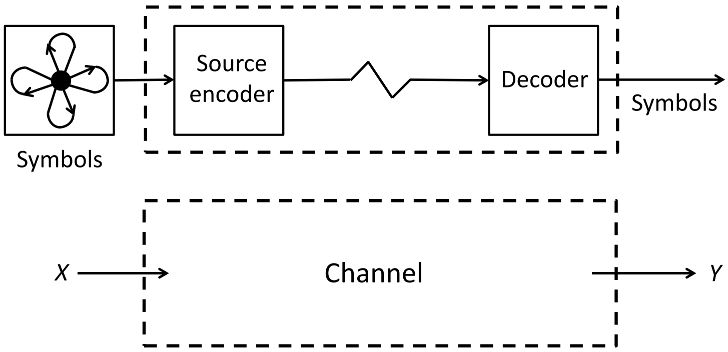
(*b*) Describe two *formative* empirical research methods that could contribute to the development of the novel wearable device, one that results in qualitative data, and one that results in quantitative data. For *each* method, describe what attributes would be expected of good quality data. [6 marks]

(*c*) Describe two *summative* empirical research methods that could contribute to the development of the novel wearable device, one that results in qualitative data, and one that results in quantitative data. For *each* method, describe what attributes would be expected of good quality data. [6 marks]

(*d*) Describe two *analytic* research methods that could be used to compare specific options in the design of the novel wearable device, one that expresses those options in qualitative terms, and one that expresses options in quantitative terms. [6 marks]

10 Information Theory

(*a*) Consider a discrete memoryless channel whose input symbol source is a random variable $X \in \{x_1, \ldots, x_J\}$ having probability distribution $p(x_j)$, and whose output symbol (possibly corrupted) is a random variable $Y \in \{y_1, \ldots, y_K\}$ (see figure below).

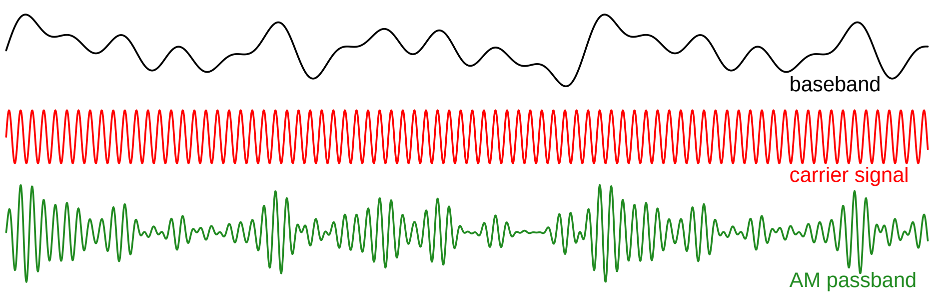


(*i*) Provide its channel matrix. [3 marks]

(*ii*) Give the average probability of correct reception, meaning the probability that the same symbol is emitted as was injected into the channel, averaged over all the cases. [3 marks]

(*b*) Show that convolution of any continuous signal with a Dirac delta function reproduces the signal. [4 marks]

(*c*) A frequency-shifting modulation of signals into different channels of a shared medium multiplies the baseband signal $f(t)$ by a complex exponential carrier wave e^{ict} of some (channel-specific) frequency c to produce a passband $f(t)\,e^{ict}$ (see figure below). Upon reception of such a passband, what process of *demodulation* would recover the original baseband?



[5 marks]

(*d*) Explain the "information diagram" of Gabor, and why the Uncertainty Principle gives it a quantal structure with an irreducible representation of the data. [5 marks]

11 Natural Language Processing

The distributional hypothesis states that the meaning of a word can be defined by its use and, therefore, it can be represented as a distribution of contexts in which the word occurs in a large text corpus.

(*a*) Describe four different types of context that can be used for this purpose. [4 marks]

(*b*) The contexts can be weighted using Pointwise Mutual Information (PMI). Explain, giving formulae, how PMI is calculated and how individual probabilities are estimated from a text corpus. [5 marks]

(*c*) Some words occur very rarely in the corpus. How does this affect their PMI scores as contexts? [4 marks]

(*d*) The goal of distributional word clustering is to obtain clusters of words with similar or related meanings. The following clusters have been produced in two different noun clustering experiments:

`Experiment 1:`

`carriage bike vehicle train truck lorry coach taxi`
`official officer inspector journalist detective constable policeman reporter`
`sister daughter parent relative lover cousin friend wife mother husband brother father`

`Experiment 2:`

`car engine petrol road driver wheel trip steering seat highway sign speed`
`concert singer stage light music show audience performance ticket`
`experiment research scientist paper result publication laboratory finding`

(*i*) How are the clusters produced in the two experiments different with respect to the similarity they capture? What lexico-semantic relations do the clusters exhibit? [3 marks]

(*ii*) The same clustering algorithm, K-means, was used in both experiments. What was different in the setup of the two experiments that resulted in the different kinds of similarity captured by the clusters? [4 marks]

12 Optimising Compilers

The following C-like program reads a `List` of integers by making calls to external procedures `readint()` to read and return an integer and `cons()` to construct a `List`:

```
    struct List { int hd; List *tl; };
    List *readlist() {
L0:   List *p = 0, *q, *t;  int v;
L1:   // comment for Part (e)
L2:   v = readint();
L3:   if (v < 0) return p;
L4:   t = cons(v, 0);
L5:   if (p == 0) {
L6:       p = t; q = t;
      } else {
L7:       q->tl = t; q = t;
      }
L8:   goto L2;
    }
```

(*a*) Define the concept of a variable being live at a program point, distinguishing a semantic notion from one which is more practically computable. Sketch an algorithm to compute the set of variables live at each program point. [5 marks]

(*b*) Give the sets of live variables your algorithm would compute for the program points `L0` ... `L8`. Also indicate any discrepancies between these and sets derived from your semantic notion of liveness. [5 marks]

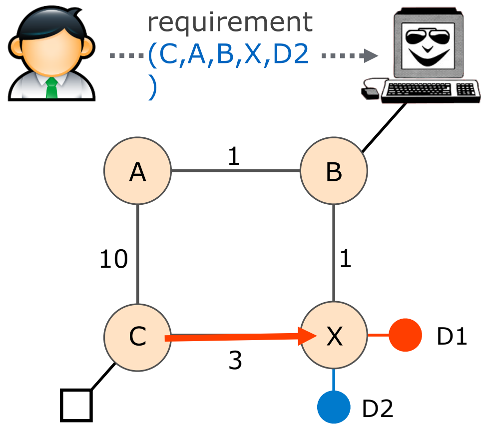
(*c*) Does the above code have any data-flow anomalies? If so explain what they are and what action a helpful compiler might take. [2 marks]

(*d*) Suppose the above code were compiled for an ARM-like processor which assumes registers 0–3 are corrupted by procedure call and registers 4–7 are preserved by procedure call; register 0 is also used for the first argument to and result from a procedure call. Assuming a register allocator operating by graph colouring, which registers might be allocated to variables `p`, `q`, `t` and `v`? Indicate when there is a choice between equivalent allocations and when there is a benefit of using one register over another. What, if anything, would be wrong with simply allocating `p`, `q`, `t` and `v` respectively to `r4`, `r5`, `r6` and `r7`? [5 marks]

(*e*) Now suppose that the commented line `L1` were replaced with some more complex code (which does not use `p`, `q`, `t` or `v`) and the register allocator found that 9 registers were now required for `readlist()` when only `r0` ... `r7` are available. How would an allocator proceed for the code as written? Can you suggest an adjustment to the source code (respecting your company's coding requirement that all declarations, but not necessarily initialisation, occur at the very start of a procedure) to allow all variables to be allocated a register? [3 marks]

13 Principles of Communications

(*a*) Distributed link-state routing algorithms can be enhanced by a central controller, such as the *fibbing* scheme, to modify the results of the computation. Explain by an example, starting from the topology below, how a specific path can be added from router C to host D2 via nodes A, B and X, by the introduction of virtual nodes in the topology, and their advertisement via the routing protocol.



[10 marks]

(*b*) Dynamic Alternative Routing (DAR), also known as Sticky Random Tandem, is a way to shed load from the most direct path in a circuit switched network to other, slightly longer paths. It depends on properties of the backbone topology of the telephone network, and, to some extent, on the properties of telephone call statistics. Describe the basic algorithm and give an overview of why it works well. [10 marks]

14 Security II

A taxpayer, whose total wealth is 300 Galactic Credits (GC), is having her tax return investigated under suspicion of irregularity. The tax inspector offers her a choice: either the investigation goes ahead as normal, which may result in her being cleared or in her having to pay a hefty fine, or she pays a fixed settlement upfront and the investigation is closed. We formalize these options respectively as the *contest* choice (40% chance of being cleared and 60% chance of paying 10 GC) and the *settle* choice (paying a fixed settlement of $s = 5$ GC). Pretend, for simplicity, that the utility of wealth in Expected Utility Theory is $U(w) = \ln(1+w)$ and that the Prospect Theory diagram has $V(x) = \ln(1 + x)$ in the first quadrant and $V(x) = -2\ln(1 + |x|)$ in the fourth quadrant, with $w, x \in \mathbb{R}$ expressed in GC.

(*a*) Answer the following two questions separately for each of the three cases of (1) straightforward probabilistic computation, (2) Expected Utility Theory and (3) Prospect Theory. Compute any numerical values to three decimal digits using a calculator.

(*i*) Will the taxpayer prefer *settle* or *contest* in each of the three cases? Justify your answer. [5 marks]

(*ii*) What should the value of s be for the taxpayer to be indifferent between the *settle* and *contest* choices in each of the three cases? [5 marks]

(*b*) For what range of values of s would the preferences of the taxpayer in the three cases of part (*a*) be *settle*, *contest*, *contest* respectively? [5 marks]

(*c*) For the case of Expected Utility Theory, draw a full-page diagram and explain in detail how to derive *graphically* the value of s for which the choice between *contest* and *settle* does not matter. [5 marks]

END OF PAPER