Specialized Subjects

10:00-12:30, Wednesday, August 20, 2025

<u>Instructions</u>

1. Do not open this booklet before the examination begins.

2. This booklet contains five problems. The number of pages is six excluding this cover sheet and blank pages. If you find missing or badly printed pages, ask the proctor for exchange.

3. Answer three problems. You can select any three out of the five. Your answer to each problem should be written on a separate sheet. You may use the reverse side of the sheet if necessary.

4. Fill the top parts of your three answer sheets as instructed below. Before submitting your answer sheets, make sure that the top parts are correctly filled.

専　門　科　目

第　　問

↑ Write the problem No.

受験番号				

↑ Write the examinee No.

5. Submit all the three answer sheets with the examinee number and the problem number, even if your answer is blank.

6. Answer either in Japanese or in English.

7. This booklet and the scratch paper must be returned at the end of the examination.

8. This English translation is supplemental and provided for the convenience of applicants. The Japanese version is the formal one.

Problem 1

Consider the circuit consisting of an operational amplifier, a resistor (resistance R_1), and a capacitor (capacitance C) as shown in Fig. 1.

(1) Let the input voltage be $v_1(t)$ and the output voltage be $v_2(t)$ at time t. Express $v_2(t)$ using $v_1(t)$. Assume that the capacitor is discharged at $t = 0$.

(2) Answer the general name of this circuit.

Consider the circuit consisting of an operational amplifier and two resistors (resistance R_2 and R_3) as shown in Fig. 2. Assume that the output voltage of this operational amplifier is restricted between $-E$ and E $(E > 0)$. Consider the voltages v_3, v_4, and v_5 shown in Fig. 2.

(3) Assume that v_3 changes from a sufficiently low voltage ($v_3 \ll 0$) to a sufficiently high voltage ($v_3 \gg 0$). Illustrate the relationship between v_3 and v_5 with v_3 on the horizontal axis and v_5 on the vertical axis. Similarly, illustrate the relationship between v_3 and v_4.

(4) Assume that v_3 changes from a sufficiently high voltage ($v_3 \gg 0$) to a sufficiently low voltage ($v_3 \ll 0$). Similarly to (3), illustrate the relationship between v_3 and v_5, and the relationship between v_3 and v_4.

As shown in Fig. 3, consider the circuit which combines the circuits shown in Fig. 1 and Fig. 2. The output voltage v_8 underwent a periodic change.

(5) Illustrate the time change of the voltages v_6, v_7, and v_8. Assume that the capacitor is discharged at $t = 0$, and the voltage v_7 is E at $t = 0$.

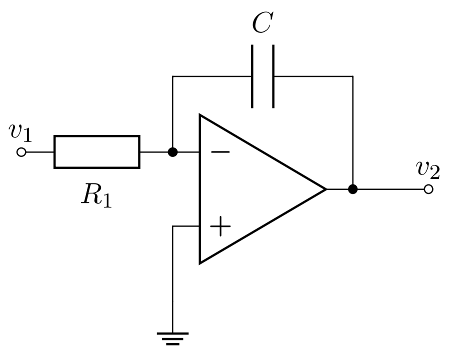

Fig. 1

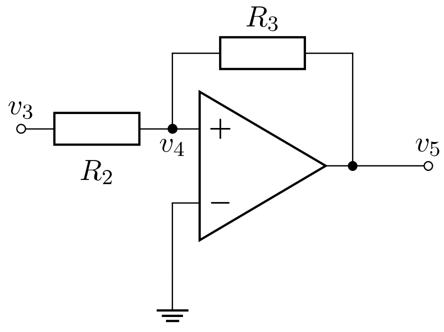

Fig. 2

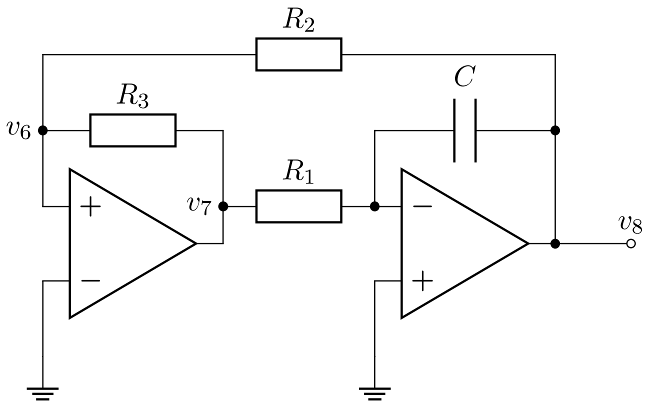

Fig. 3

Problem 2

Consider an in-order execution processor with a six-stage pipeline stage structure as shown in the figure below. Assume a simple RISC-type architecture as the instruction set architecture. Answer the following questions.

Fetch Stage	Decode Stage	Register Read Stage	Execution Stage	Memory Stage	Register Write Stage

(1) Describe the operations performed and the hardware used for each stage from Fetch to Register-write as they occur in order. Use all of the following terms appropriately and underline where they are used.
register file, main memory, instruction cache, data cache,
ALU, operand, opcode

(2) Assume the delays required for processing each stage as shown in the table below. In addition, assume that the overhead caused by the pipeline register is 200 ps for each stage. Given these conditions, indicate the maximum operating frequency of this processor.

Fetch	400 ps
Decode	100 ps
Register Read	200 ps
Execution	300 ps
Memory	600 ps
Register Write	200 ps

(3) Given the conditions and the maximum operating frequency in (2), show the time required for a single instruction to be processed from start to completion in this processor. In addition, show the expected processing time for a single instruction when this processor is not pipelined, and describe why pipelining is expected to improve the performance.

(4) Describe when data hazards related to register dependencies occur in this processor. In addition, show one appropriate countermeasure and explain it.

(5) Consider executing a program on this processor such that CPI (Cycles Per Instruction) would be 1 if branch prediction were always accurate. Answer the CPI when the branch instruction ratio is 10%, the branch prediction accuracy is 95%, and the branch prediction miss penalty is 3 cycles. The number of instructions in the program can be considered sufficiently large.

Problem 3

Given set $\mathcal{V}$ of integer arrays and integer array $S(\notin \mathcal{V})$ of length $n(\geq 2)$, consider whether S can be split into subarrays, each of which is contained in $\mathcal{V}$. For instance, for $\mathcal{V} = \{[1], [0,0], [0,1]\}$, $S = [0,1,0,0]$ can be split into $[0,1]$ and $[0,0]$. Answer the following questions.

(1) Express the number of combinations that split S into multiple subarrays, using n.

(2) Explain which data structure should be used to store $\mathcal{V}$, with a concrete implementation, to verify whether each subarray in S belongs to $\mathcal{V}$ in time independent of $\mathcal{V}$'s cardinality.

(3) The pseudocode below verifies whether S can be split into subarrays, each contained in $\mathcal{V}$.

```
find_segmentation(V, S, n):
  is_split_before[0] ← True
  for i ← 1 to n do
    is_split_before[i] ← False
  for i ← 0 to n − 1 do
    ┌──────────────┐
    │     (P)      │
    └──────────────┘
  return is_split_before[n]
```

Fill `(P)` to complete this pseudocode. You may use function `contain`$(\mathcal{V}, S, i, j)$ to verify whether $\mathcal{V}$ contains subarray $[S[i], S[i+1], \ldots, S[j]]$ in the worst time complexity of $\mathcal{O}(1)$.

(4) Show the worst time complexity of the pseudocode in (3).

(5) Modify the pseudocode in (3) to find splits of S into the smallest number of subarrays, each contained in $\mathcal{V}$. You may return a data structure that can reconstruct all valid splits.

Problem 4

Loss functions are employed to quantitatively evaluate the difference between the predicted value of a machine learning model and the ground truth value. Since many types of loss functions have been proposed, consider which types of data each loss function is effective for learning from. Given the n elements of two-dimensional data $\{(x_1, y_1), \cdots, (x_n, y_n)\}$, assume that y_i is modeled by $f(x_i) + \epsilon_i$. The function f denotes the polynomial function ($f : \mathbb{R} \to \mathbb{R}$) and ϵ_i denotes the error. Answer the following questions.

(1) When the random variable z follows a normal distribution with mean μ and variance σ^2, the probability density function of $p(z)$ is expressed as follows. When the error follows a normal distribution with mean 0 and variance σ^2, express the conditional probability density function $p(y|x)$ using y, $f(x)$, and σ.

$$p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(z-\mu)^2}{2\sigma^2}\Big).$$

(2) Consider using maximum likelihood estimation to find the parameters of $f(x)$ (for example, the parameters of the linear function $f(x) = ax + b$ are a and b) that maximize the likelihood function below. Through this calculation, explain the reason why the loss function represented by the Mean Squared Error (MSE) is effective for learning from data where the errors follow a normal distribution with mean 0.

$$\prod_{i=1}^{n} p(y_i|x_i).$$

(3) When the random variable z follows a Laplace distribution with mean μ and variance $2b^2$, the probability density function $p(z)$ is expressed as follows. Explain the reason why the loss function represented by the Mean Absolute Error (MAE) is effective for learning from data where the errors follow a Laplace distribution with mean 0.

$$p(z) = \frac{1}{2b} \exp\Big(-\frac{|z-\mu|}{b}\Big).$$

(4) MSE is easily affected by outliers, whereas MAE is robust to them. Describe its reason from the difference between the MSE and MAE equations.

(5) MAE has the advantage described in (4), but it is difficult to conduct fine-grained optimization because the $\mathcal{L}_1$ loss function expressed as $\mathcal{L}_1(y, f(x)) = |y - f(x)|$ in MAE is not differentiable when $\mathcal{L}_1(y, f(x)) = 0$ and its slope is constant even when $\mathcal{L}_1(y, f(x))$ is close to 0. To mitigate both MSE and MAE problems, the following Huber loss function has been proposed, where δ denotes a constant value. Explain how the Huber loss function mitigates both MSE and MAE problems from the viewpoint of the following equation.

$$\mathcal{L}_{Huber}(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & (|y - f(x)| \leq \delta) \\ \delta(|y - f(x)| - \frac{1}{2}\delta) & (\text{otherwise}). \end{cases}$$

Problem 5

Answer the following questions. Use $\log_2 3 = 1.6$ and $\log_2 5 = 2.3$ in your calculations.

Consider memoryless, stationary, quaternary source S that generates source symbols A, B, C, and D with probabilities $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{8}$, and $\frac{3}{8}$, respectively.

(1) Compute the entropy $H(S)$ of S.

(2) Show the process of constructing a binary Huffman code for S. Also, compute the expected codeword length L per source symbol.

(3) When each source symbol of S is encoded into a uniquely decodable binary code, briefly explain the relationship that holds between $H(S)$ and L.

Consider memoryless, stationary, binary symmetric channel (BSC) W with crossover probability p.

(4) Express the channel capacity of W in terms of p.

(5) Assuming that $p = 0.2$ in W, consider channel coding using a simple repetition code in which each symbol is sent three times (symbol 0 is sent as 000; symbol 1 is sent as 111) and decoded by majority voting. Then, compute (a) the channel capacity, (b) the information transmission rate, and (c) the probability of decoding error per codeword.

(6) If the information transmission rate R is below the channel capacity C of BSC, briefly explain what kinds of codes are theoretically achievable, based on the channel coding theorem.

(7) Derive the following. Assume that neither encoding nor decoding is performed between cascaded BSCs.

- (a) The crossover probability of the channel obtained by concatenating two instances of W in series is $2p - 2p^2$.
- (b) The crossover probability of the channel obtained by concatenating n instances of W in series is $\frac{1}{2}\left(1 - (1 - 2p)^n\right)$.