
Examiners' commentaries

2016–17

CO3325 Data compression – Zones A and B

General remarks

Data compression is a technical subject and the examination questions are precise and specific. Each question consists of a number of parts, and each part may contain a number of implicit or explicit sub-questions. To achieve a good grade, you must ensure that you read the questions carefully and understand what is required, taking care to check whether there are any implicit tasks involved.

The examination can be viewed as a one-off written communication between you and the examiner. As it gives no second chance for clarification and is highly constrained by time, you should aim to demonstrate your knowledge of the subject in the most efficient manner in the examination. For example, you should consider whether it would be better to draw a diagram with annotation and explanation for certain questions, or to provide a detailed written response.

As Data compression is a Level 6 examination, you will need to demonstrate your competence in problem solving. This sometimes includes sharing your interpretation of given specifics in questions. For example, you may add a note on your script to clarify such issues in order to avoid any potential confusion, or use notes to add information about assumptions you have made in order to justify your solutions.

The parts and sub-questions in the examination do not necessarily have equal levels of complexity. You should try to secure as many marks as possible as early as possible. For example, you may wish to tackle questions you are most comfortable with first, or, where you are less sure of the answers, give only itemised answers and add details later if time permits. In such cases, please make sure that you read the whole question, to understand the scope of the answers expected in each part. Where possible, leave space in your answer books for anything you may wish to add if time permits.

If you follow these guidelines, as well as those in the subject guide and textbooks, your examination should be an enjoyable experience. It is a good opportunity for you to check the level of your knowledge and to celebrate your academic achievements.

Comments on specific questions – Zone A

Question 1

- a. This part of the question addressed the absolute limit of lossless data compression. There are two implicit sub-questions and both should be answered. The most common error was failure to read the question carefully. Some candidates omitted to give explanation or an example, or did not provide a percentage for the proportion of files that can be compressed by one byte.
- b. This part was straightforward, requiring an explanation and an example. Some candidates gave only a brief definition; others gave flawed explanations, showing poor knowledge of the topic.

- c. There were two implicit sub-questions in this part. However, some candidates answered only one: for example, they explained why the Reflected Gray code is good for greyscale images but forgot to derive the code; or vice versa: they gave no explanation, but provided the code. A few candidates gave incorrect explanations, showing a knowledge gap.
- d. This part contained two explicit sub-questions. When writing your answers in the examination, please make sure that the sub-question numbers are used correctly and that answers are full and complete. Most candidates explained the concepts correctly, but the specific differences were less well handled.

Question 2

- a. There were two implicit sub-questions in this part. The most common error was to have answered only one of the two sub-questions. For example, the problem was described but no solution was provided. Another mistake was to confuse the efficiency problem with the effectiveness problem. Please note that the efficiency problem can be addressed by maintaining two lists.
- b. This part of the question was straightforward and was well answered in general. Good answers included a sequence of iterative steps showing the values for the required variables in the decoding process. Some candidates lost marks unnecessarily, as they failed to give each step in the process, as specified by the question.
- c. There were three explicit sub-questions in this part.
 - i. This required an extended alphabet, and was well answered in general.
 - ii. This required a probability distribution, also well answered by most candidates.
 - iii. This required demonstration that, for any probability distribution, the (first-order) entropy value of the extended alphabet is doubled. Some candidates showed the doubling relationship for the given probability case in values computed in previous sub-questions, missing the fact that the sub-question asks candidates to show the relationship 'for any P_i ' cases.

Question 3

- a. This part of the question was straightforward, with two implicit sub-questions which were generally well answered. Some candidates were unable to provide correct explanations of what is meant by 'sampling' in data compression – a basic topic.
- b. There were two explicit sub-questions in this part. Some candidates only answered one sub-part and so lost marks.
 - i. Some candidates misunderstood the relationship between the Kraft-McMillan inequality and a uniquely decodable code. Please note that it is possible to find a uniquely decodable code of the given length distribution if the inequality is satisfied. However, a code of the given length distribution is not necessarily uniquely decodable. Similarly, a prefix code is always uniquely decodable, but a uniquely decodable code is not necessarily a prefix code.
 - ii. This sub-part was well answered by most candidates.
 - iii. This part also contained two explicit sub-questions. A common error made by some candidates this year was to give only partial results. For example, they gave the values of the input and the tree, but did not provide the values for other required parts such as those of output and alphabet, and so lost marks.

Comments on specific questions – Zone B

Question 1

- a. There were two implicit sub-questions in this part. The second sub-question was better answered than the first. Most candidates could convert the given decimal number to the Reflected Gray code. However, some incorrect answers were given for the first sub-question. For example, that RG code could be presented in binary code and separated into bitplanes for further individual compression when in fact normal binary codes can also be separated into bitplanes.
- b. This part of the question was well answered.
- c. There were three explicit sub-questions in this part.
 - i. This was straightforward, and generally answered well, but some candidates failed to provide the flowchart or pseudocode as required.
 - ii. This was again well-answered, with correctly completed tables, as required.
 - iii. This simply required the decoded string output. Some candidates failed to provide this final value and lost marks.

Question 2

- a. This part of the question was straightforward and was generally well answered. Unfortunately, some candidates did not provide the Huffman code as required. Some others failed to provide trees.
- b. Again, this part was generally answered well. Most candidates were able to show the decoding steps, and sensibly opted to provide a table showing the values of required variables for each step. Unfortunately, some candidates did not give the final answer.
- c. This part of the question included three explicit sub-questions.
 - i. This was straightforward, and answered well, simply requiring an extended alphabet.
 - ii. This was again well answered, although some candidates made errors in logarithm calculations.
 - iii. This sub-part was not attempted by many candidates, although it is not very difficult to answer. The critical step is to realise the fact that $p_{AB} = p_A p_B$ and $\log AB = \log A + \log B$ and $p_A + p_B = 1$. Some candidates used the computation results in their answers to previous sub-questions to show the relationship between the quantities of the two entropy values. That is flawed, unfortunately, as it can demonstrate the relationship for one case, but not for cases of **any** probability distribution P , as required.

Question 3

- a. This part of the question was straightforward, and required both an explanation and an example, supported by simple colour map. Unfortunately, some candidates were insufficiently familiar with the topic to answer this question well.
- b. There were three implicit sub-questions in this part, which was well answered in general. Some confusions were apparent: for example, confusing a 'code' with a 'codeword', and a 'code with the same lengths as those of the given code' and 'the given code'. Please make sure that you read the question carefully, to ensure you understand what is being asked.
- c. This part was straightforward and was generally well answered where attempted, with either a flowchart or pseudocode. Unfortunately, many candidates did not attempt it, showing a knowledge gap. Furthermore, some candidates did not give a flowchart or pseudocode as required, but

provided a full example (which was actually more time-consuming) for which some credit was given.

- d. There were two explicit sub-questions in this part.
 - i. This was well answered, with each value clearly shown.
 - ii. This was poorly answered, suggesting a knowledge gap for 'compression factor'.