ERROR CORRECTION, COUNTABILITY 5

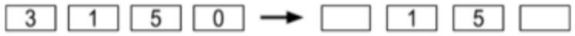
COMPUTER SCIENCE MENTORS 70

October 1 - 5, 2018

1 Erasure Errors

1.1 Introduction

We want to send n packets and we know that k packets could get lost. We use a polynomial under GF(q).



How many more points does Alice need to send to account for k possible errors? __

Solution: k

What degree will the resulting polynomial be? ___

Solution: $n - 1$

How large should q be if Alice is sending n packets with k erasure errors, where each packet is a integer between 0 and m?

Solution: Modulus should be larger than $n + k$, larger than m, and prime

What would happen if Alice instead sends $n + k - 1$ points? Why will Bob be unable to recover the message?

Solution: Bob will receive $n - 1$ distinct points and needs to reconstruct a polynomial of degree $n - 1$; this is impossible. There are q polynomials of at most degree $n - 1$ in $GF(q)$ that go through the $n - 1$ points that Alice sent.

1.2 Questions

1. Suppose $A = 1$, $B = 2$, $C = 3$, $D = 4$, and $E = 5$. Assume we want to send a message for which the information is contained in 3 packets. Recover the lost part of the message, or explain why it cannot be done.
   1. C_AA

      Solution: $P(0) = 3, P(2) = 1, P(3) = 1$. q must be greater than m, which is 5, and $n + k$, which is 4, and also prime. The smallest number that satisfies these conditions is 7, so we interpolate the polynomial over $\mod 7$ and get

$5x^2 + 3x + 3$. Now, once we evaluate this at 1, we get 4. So, in the end, it's CDAA.

2. CE_ _

Solution: Impossible. In order to get the original degree 2 polynomial, we need at least 3 points.

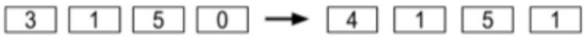
2. Suppose we want to send n packets, and we know $p = 20\%$ of the packets will be erased. How many extra packets should we send? What happens if p increases (say to 90%)?

Solution: We want to have $(1-p)*(n+k) = n$, where k is the number of additional packets we send. This is because the fraction of packets not erased * number of packets that are sent = number of packets in original message. Solving for k, we get $\frac{n}{1-p} - n$. When p is large, we have to send many times the number of original packets.

2 General Errors (Berlekamp and Welch)

2.1 Introduction

Now instead of losing packets, we know that k packets are corrupted. Furthermore, we do not know which k packets are changed. Instead of sending k additional packets, we will send an additional $2k$. This ensures that we have at least $n + k$ un-corrupted points and can uniquely determine the original polynomial.



Solomon-Reed Codes

1. Identical to erasure errors: Alice creates $n - 1$ degree polynomial $P(x)$.

$$P(x) = p_0 + p_1x + \ldots + p_{n-1}x^{n-1}$$

2. Alice sends $P(1), \ldots, P(n + 2k)$

3. Bob receives $R(1), \ldots, R(n + 2k)$

For how many points does $R(x) = P(x)$?

Solution: $n + k$

True or false: $P(x)$ is the unique degree $n - 1$ polynomial that goes through at least $n + k$ of the received points.

Solution: True

Berlekamp Welch

How do we find the original polynomial $P(x)$? Suppose that $m_1, \ldots, m_k$ are the corrupted packets. Let $E(x) = (x - m_1) \ldots (x - m_k)$. Then $P(i) * E(i) = r_i * E(i)$ for any i greater than 1 and less than $n + 2k$. Why?

Solution: Case 1: i is corrupted

$$E(i) = 0 \rightarrow 0 = 0$$

Case 2: i is not corrupted

$$P(i) = r_i \rightarrow P(i) * E(i) = r_i E(i)$$

Let $Q(i) = P(i)E(i)$. So we have $Q(i) = P(i)E(i) = r_i * E(i)$ where $1 \leq i \leq 2k + n$. What degree is $Q(i)$?

Solution: $n + k - 1$

How many coefficients do we need to describe $Q(i)$?

Solution: $n + k$

What degree is $P(i)$?

Solution: $n - 1$

How many unknown coefficients do we need to describe $E(i)$?

Solution: k (we know that the first coefficient has to be 1)

We can write $Q(i) = r_i E(i)$ for every i that is $1 \leq i \leq 2k + n$.
How many equations do we have? How many unknowns?

Solution: $n + 2k$

Once we have the above described equations, how do we determine what $P(i)$ is?

Solution: Solve the equations to get the coefficients for $E(i)$ and $Q(i)$. Then divide $\frac{Q(i)}{E(i)}$ to get $P(i)$.

2.2 Questions

1. (a) Alice sends Bob a message of length 3 on the Galois Field of 5 (modular space of mod 5). Bob receives the following message: (3, 2, 1, 1, 1). Assuming that Alice is sending messages using the proper general error message sending scheme, set up the linear equations that, when solved, give you the $Q(x)$ and $E(x)$ needed to find the original $P(x)$.

Solution: To determine $Q(x)$ and $E(x)$, set up 5 equations for the five values of x that we have, such that

$$Q(x) = r_x * (x - b) \mod 5$$

where

$$Q(x) = a_3 * x_i^3 + a_2 * x_i^2 + a_1 * x_i + a_0$$

(r_i = ith received number)
$x = 1 : a_3 + a_2 + a_1 + a_0 = 3(1 - b)$ mod 5
$x = 2 : 3a_3 + 4a_2 + 2a_1 + a_0 = 2(2 - b)$ mod 5
$x = 3 : 2a_3 + 4a_2 + 3a_1 + a_0 = 1(3 - b)$ mod 5
$x = 4 : 4a_3 + 1a_2 + 4a_1 + a_0 = 1(4 - b)$ mod 5
$x = 5 : 0 + 0 + 0 + a_0 = 1(0 - b)$ mod 5

The solution to these equations is: $b = 3, a_3 = 1, a_2 = 3, a_1 = 3, a_0 = 2$.

(b) What is the encoded message that Alice actually sent? What was the original message? Which packet(s) were corrupted?

Solution: This means that $Q(x) = x^3 + 3x^2 + 3x + 2$ and $E(x) = x - 3$.
To find the actual encoded message we use $P(x) = \frac{Q(x)}{E(x)}$. We divide and find that $P(x) = x^2 + 6x + 21 + \frac{65}{x-3} = x^2 + x + 1$ mod 5. We plug in the values 1, 2, 3, 4, 5 to find the encoded 5 packet message.
$P(1) = 1^2 + 1 + 1 = 3$ mod 5
$P(2) = 2^2 + 2 + 1 = 7 = 2$ mod 5
$P(3) = 3^2 + 3 + 1 = 13 = 3$ mod 5
$P(4) = 4^2 + 4 + 1 = 21 = 1$ mod 5
$P(5) = 5^2 + 5 + 1 = 31 = 1$ mod 5
The actual message is (3,2,3,1,1). We see that the 3rd packet doesn't match the initial message of (3,2,1,1,1), and is therefore the corrupted one (we could also have seen this by looking at the value of b.

2. You want to send a super secret message consisting of 10 packets to the space station through some astronauts. You are afraid that some malicious spy people are going to tell the wrong message and make the space station go spiraling out of orbit. Assuming that up to 5 of the astronauts are malicious, design a scheme so that the group of astronauts (including the malicious ones) still find the correct message that you want to send. You can send any number of astronauts, but try to make the number that you have to send as small as possible. Astronauts can only carry one packet with them.

Solution: We can use the scheme provided in the notes for general errors. If there are up to 5 malicious astronauts, then we have up to 5 general errors in our transmission. To account for those errors and to send the original message we have to send at least the length of the original message plus the twice the possible number of general errors. Thus, we have to send 20 astronauts or more. Create a polynomial of degree 9 (which is the length of the message minus 1), such that $P(0), P(1), \ldots, P(10)$ is the original message, and give each of the 20 astronauts unique points on that polynomial.

3 Countability

3.1 Introduction

(a.) Cardinality is defined as the number of elements in a set. We define a set as countably infinite if it has the same cardinality as the natural numbers (or any countable set).
(b.) We can prove a set is countable by finding a bijection between the set and any countable set. A few classic examples are the hotel argument to show that $\mathbb{Z}^+$ is countable, and the spiral argument to show that the $\mathbb{Q}$ is countable, both included in your notes.
(c.) To prove a set is uncountable, we can either find a bijection between it and an uncountable set or use the Cantor Diagonalization proof, included in your notes.

3.2 Questions

1. **True/False**
   (a) Every infinite subset of a countable set is countable

Solution: True. Define set S as the infinite subset, and set T as the countable set. $S \subseteq T$, which means $|S| \leq |T|$. This means S is countable. S is also an infinite subset, so it is infinite. Therefore, S is countably infinite.

(b) If A and B are both countable, then $A \times B$ is countable

Solution: True. Can draw a bijection where first elem of A maps to 1, first elem of B maps to 2, second elem of A maps to 3, etc. This will include $A \times B$ at the end, and because there is a bijection from A to $\mathbb{N}$ and B to $\mathbb{N}$, there is a bijection from $A \times B$ to $\mathbb{N}$. There is clearly a mapping from $A \times B$ to $\mathbb{N} \times \mathbb{N}$, and $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$, so there is a bijective mapping from $A \times B$ to $\mathbb{N}$.

(c) If A_i is countable, then $A_1 \times A_2 \times A_3 .... \times A_N$ for N finite is countable.

Solution: We can give a kind of a "collapsing" argument. We know that $A_1 \times A_2$ is countable, so we can just define $A_{12} = A_1 \times A_2$, then we have $A_{12} \times A_3$.. and so on.

(d) If A_i is countable, then $A_1 \times A_2 \times A_3 ...$ a countably infinite number of times is countable.

Solution: False. Consider the real numbers - each "digit" is picked from a finite countable set, crossed a countable number of times, but the entire set is uncountable.

(e) Every infinite set that contains an uncountable set is uncountable.

Solution: True. Let A be an uncountable subset of B. Assume that B is countable. $f : \mathbb{N} \to B$ is a bijection. There must be a subset M of $\mathbb{N}$ such that $f : M \to A$ is a bijection. Then A is countable. This is a contradiction. So B must be uncountable.

2. Are these sets countably infinite/uncountably infinite/finite? If finite, what is the order of the set?

(a) Finite bit strings of length n.

Solution: Finite. There are 2 choices (0 or 1) for each bit, and n bits, so there are $2 \times 2 \times \ldots \times 2 = 2^n$ such bit strings.

(b) All finite bit strings of length 1 to n.

> **Solution:** Finite. By part (a), there are 2^1 bit strings of length 1, 2^2 of length 2, etc. Thus, there are $2^1 + 2^2 + \ldots + 2^n = 2^{n+1} - 2$.

(c) All finite bit strings

> **Solution:** Countably infinite. We can list these strings as follows: $\{0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111, 0000, \ldots\}$. This gives us a bijection with the (countable) natural numbers, so these are countably infinite.

(d) All infinite bit strings

> **Solution:** Uncountably infinite. We can construct a bijection between this set and the set of real numbers between 0 and 1. We can represent these real numbers using binary e.g. they are of the form $0.0110001010110\ldots$. By diagonalization, the set of real numbers between 0 and 1 is uncountably infinite; therefore, so is this set.

(e) All finite or infinite bit strings.

> **Solution:** Uncountably infinite. This is the union of a countably infinite set (part c) and an uncountably infinite set (part d), so it is uncountably infinite.

3. Find a bijection between N and the set of all integers congruent to $1 \mod n$, for a fixed n.

> **Solution:** The set of integers congruent to 1 mod n is $A = \{1 + kn \mid k \in \mathbb{Z}\}$. Define $g : \mathbb{Z} \to A$ by $g(x) = 1 + x \times n$; this is a bijection because it is clearly one-to-one, and is onto by the definition of A. We can combine this with the bijective mapping $f : \mathbb{N} \to \mathbb{Z}$ from the notes, defined by $f(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ \frac{-(x+1)}{2} & \text{if } x \text{ is odd} \end{cases}$. Then $f \circ g$ is a function from $\mathbb{N}$ to A, which is a bijection.

4. Are the power sets S of a countably infinite set are finite, countably infinite, or uncountably infinite? Provide a proof for your answer.

Solution: The power sets of a countably infinite set are uncountably infinite. There is a bijection between the set $2^{\mathbb{N}}$ and 2^S, as S and $\mathbb{N}$ have the same cardinality. The set $2^{\mathbb{N}}$ is uncountable. We prove this through contradiction. We assume the set $2^{\mathbb{N}}$ is countably infinite. This means we can list the subsets of $\mathbb{N}$ such that every subset is N_i for some i. We define another set $A = \{i | i \geq 0 \text{ and } i \notin N_i\}$ which contains integers i not part of N_i. But, N is a subset of $\mathbb{N}$ so we must have $N = N_j$ for some j. This means that if $j \in N$, then $j \notin N$, and if $j \notin N$, then $j \in N$. This is a contradiction since j is either in N or not, so the set is not countably infinite.