1. Problem Set

(1) True or False
  (a) There are finitely many finite words on the finite alphabet $\{0, 1\}$.
  (b) The Thue-Morse word is square-free.
  (c) The Thue-Morse word is cube-free.
  (d) The Thue-Morse words is quartic-free (i.e. No subwords of the form u^4, where u is a nonempty word.)
  (e) Every infinite word has a unique prefix. (i.e. only one prefix)
  (f) The word 010330023023011230 is overlap-free.
  (g) There exists a square-free word on the alphabet $\{0, 1, 2, 3\}$.
  (h) $t_p = t_q$ for all prime numbers $p, q \in \mathbb{N}$.

(2) Prove that if w is a word of length n then w has at most $\dfrac{n^2+n+2}{2}$ subwords. (Hint: $1 + 2 + 3 + \cdots + n = \dfrac{n(n+1)}{2}$)

(3) Give an example of a word of length 4 over the alphabet $\{0, 1, 2, 3, 4, 5\}$ which has $\frac{16+4+2}{2} = 11$ subwords. This shows that the bound given in problem (2) may be actually achieved by some word.

(4) How many words on the alphabet of $\{0, 1\}$ of length 7 are there?

(5) Determine t_n for $n = 65, 130$, and $2^{1000000} + 3$.

(6) Prove that if w is a subword of t then $\overline{w}$ is a subword of t.

(7) Write down the first 10 letters of c.

(8) Show that 010 is not a subword of c.

(9) Consider

$$q = (q_n)_{n=0}^{\infty} = 11001000010000001\cdots$$

Predict a way to determine q_n, given any $n \in \mathbb{N}$. Be sure to check your answer with me as you will be using this sequence for the following problems

(10) Prove that q contains a word of the form u^k, where $u \neq \varepsilon$, for every $k \in \mathbb{N}$.

(11) (Don't spend too much time on this question :))Write down a word (of length at least 3) in the English language which is
  (a) square-free
  (b) cube-free but not square-free
  (c) a square
  (d) an overlap (find one other than alfalfa!)

(1) True or False
  (a) There are finitely many finite words on the finite alphabet $\{0, 1\}$.
  (b) The Thue-Morse word is square-free.
  (c) The Thue-Morse word is cube-free.
  (d) The Thue-Morse words is quartic-free (i.e. No subwords of the form u^4, where u is a nonempty word.)
  (e) Every infinite word has a unique prefix. (i.e. only one prefix)
  (f) The word 010330023023011230 is overlap-free.
  (g) There exists a square-free word on the alphabet $\{0, 1, 2, 3\}$.
  (h) $t_p = t_q$ for all prime numbers $p, q \in \mathbb{N}$.

(2) Prove that if w is a word of length n then w has at most $\dfrac{n^2+n+2}{2}$ subwords. (Hint: $1+2+3+\cdots+n=\dfrac{n(n+1)}{2}$)
(3) Give an example of a word of length 4 over the alphabet $\{0,1,2,3,4,5\}$ which has $\frac{16+4+2}{2}=11$ subwords. This shows that the bound given in problem (2) may be actually achieved by some word.
(4) How many words on the alphabet of $\{0,1\}$ of length 7 are there?
(5) Determine t_n for $n=65, 130$, and $2^{1000000}+3$.
(6) Prove that if w is a subword of t then $\overline{w}$ is a subword of t.
(7) Write down the first 10 letters of c.
(8) Show that 010 is not a subword of c.
(9) Consider

$$q=(q_n)_{n=0}^{\infty}=11001000010000001\cdots$$

Predict a way to determine q_n, given any $n \in \mathbb{N}$. Be sure to check your answer with me as you will be using this sequence for the following problems
(10) Prove that q contains a word of the form u^k, where $u \neq \varepsilon$, for every $k \in \mathbb{N}$.
(11) (Don't spend too much time on this question :))Write down a word (of length at least 3) in the English language which is
   (a) square-free
   (b) cube-free but not square-free
   (c) a square
   (d) an overlap (find one other than alfalfa!)

2. Solutions

(1) True or False

(a) There are finitely many finite words on the finite alphabet $\{0, 1\}$.

False. Notice that we place no restriction on the length of such a word. Therefore, in particular, $1, 11, 111, 1111, \ldots$ is an infinite list of finite words on the alphabet $\{0, 1\}$.

(b) The Thue-Morse word is square-free.

False. We proved that c is square-free, not t. In particular, 11 is a subword of t which is a square.

(c) The Thue-Morse word is cube-free.

True. Every cube is an easy example of an overlap and t is overlap-free.

(d) The Thue-Morse words is quartic-free (i.e. No subwords of the form u^4, where u is a nonempty word.)

True. If t contained a quartic then it would contain a cube, which is impossible by the above question.

(e) Every infinite word has a unique prefix. (i.e. only one prefix)

False. Every infinite word has a unique prefix of a given length. Therefore every infinite word has infinitely many prefixes.

(f) The word 010330023023011230 is overlap-free.

False. We see that 0230230 is a subword which is an overlap. Namely, we take $c = 0$ and $x = 23$.

(g) There exists a square-free word on the alphabet $\{0, 1, 2, 3\}$.

True. All you have to do is consider c over the alphabet $\{0, 1, 2, 3\}$ with no occurrences of the letter 3.

(h) $t_p = t_q$ for all prime numbers $p, q \in \mathbb{N}$.

False. Observe that $[5]_2 = 101$ and $[7]_2 = 111$ so that $t_5 = 0$ but $t_7 = 1$.

(2) Prove that if w is a word of length n then w has at most $\dfrac{n^2+n+2}{2}$ subwords. (Hint: $1+2+3+\cdots+n = \dfrac{n(n+1)}{2}$)

Let w be a word of length n. Note that w can have at most n subwords of length 1. Namely, this situation occurs when no letter occurring in w is repeated. Next, if you are to consider all the subwords of length 2 from left to right, we see that we have considered $n-1$ subwords. Therefore the number of distinct subwords we have considered is at most $n-1$. Similarly, we see that the number of subwords of w of length k is $n-k+1$ until we observe that there is at most one subword of w of length n. Therefore we have accounted for at most

$$n+(n-1)+(n-2)+\cdots+3+2+1 = \frac{n(n+1)}{2} = \frac{n^2+n}{2},$$

subwords. Notice that we are one off our target value! What did we miss? We accounted for all words of positive length! That's right! The empty word, ε, is a subword of every word. Therefore we get one more subword of w and so we have proved our claim.

(3) Give an example of a word of length 4 over the alphabet $\{0,1,2,3,4,5\}$ which has $\frac{16+4+2}{2} = 11$ subwords. This shows that the bound given in problem (2) may be actually achieved by some word.

Consider the word $w = 1234$. The subwords of this word are

$$\{\varepsilon, 1, 2, 3, 4, 12, 23, 34, 123, 234, 1234\}$$

so that there are indeed 11 subwords of w.

(4) How many words on the alphabet of $\{0,1\}$ of length 7 are there?

Each time we pick a letter to form a word of length 7 we have two choices. Namely, we may choose 0 or 1. Since we must pick 7 letters to form such a word, there are 2^7 words of length 7 over the alphabet $\{0,1\}$.

(5) Determine t_n for $n = 65, 130$, and $2^{1000000}+3$.

First note that $65 = 2^6+1$ so that $[65]_2 = 1000001$. Hence $t_{65} = 0$. Now, $130 = 2^7+2$ so that $[130]_2 = 10000010$. Hence $t_{130} = 0$. Finally, observe that $[2^{1000000}+3]_2 = 10^{999998}11$ so that $t_{2^{1000000}+3} = 1$.

(6) Prove that if w is a subword of t then $\overline{w}$ is a subword of t.

Suppose w is a subword of t. Then w is a subword of X_n for some n big enough. But then $X_{n+1} = X_n\overline{X_n}$ is also a subword of t. Since $\overline{w}$ is a subword of X_{n+1} we see that $\overline{w}$ is a subword of t.

(7) Write down the first 10 letters of c.

We first notice that

$$t = 0110100110010110100101100\cdots$$

so that now it is just a matter of counting the number of 1's between 0's. We see that

$$c = 2102012101\cdots$$

(8) Show that 010 is not a subword of c.

Well, let us see what would happen if 010 was a subword of c. Note that 0100 and 0101 are not subwords of c because they contain squares and c is square-free. Therefore, if 010 is a subword of c then 0102 is a subword of c. However, this then means that 00100110 is a subword of t. But this means that 100100110 is a subword of t since t is cube (overlap)-free. Now, this means that 1(00)1(00)1 is an overlap which is a subword of t, which is impossible! Therefore 010 is not a subword of c.

(9) Consider

$$q = (q_n)_{n=0}^{\infty} = 11001000010000001\cdots$$

Predict a way to determine q_n, given any $n \in \mathbb{N}$. Be sure to check your answer with me as you will be using this word in the next problem.
Note that $q_n = 1$ when $n \in \{0, 1, 4, 9, \ldots, 16\}$. What do these have in common? Well, they are all perfect squares. Therefore we define q_n to be 1 if n is a perfect square and 0 otherwise.

(10) Prove that q contains a word of the form u^k, where $u \neq \varepsilon$, for every $k \in \mathbb{N}$.

We claim that we can take $u = 0$. Note that to show 0^k is a subword of q for every $k \in \mathbb{N}$ it is enough to show that consecutive perfect squares get further and further apart. That is, there are more and more zeroes between the occurrences of 1 in q. To see this observe that for any $n \in \mathbb{N}$,

$$(n+1)^2 - n^2 = n^2 + 2n + 1 - n^2 = 2n + 1.$$

Therefore, between the nth and (n+1)st occurence of 1 in q there are $2n$ 0's in q.
So now take a random $k \in \mathbb{N}$. From the above work we did we know that there are $2k$ 0's between the kth and $(k+1)st$ occurence of 1 in q. Therefore 0^{2k} is a subword of q and so, in particular, 0^k is a subword of q.

(11) (Don't spend too much time on this question :))Write down a word (of length at least 3) in the English language which is

(a) square-free: cat
(b) cube-free but not square-free: silly
(c) a square: hotshots
(d) an overlap (find one other than alfalfa!): entente