

4. **Problem:** Show that any alphabet  $\Sigma$  with at least two symbols is comparable to the binary alphabet  $\Gamma = \{0, 1\}$ , in the sense that strings over  $\Sigma$  can be easily encoded into strings over  $\Gamma$  and vice versa. How much can the length of a string change in your encoding? (I.e., if the length of a string  $w \in \Sigma^*$  is  $|w| = n$  symbols, what is the length of the corresponding string  $w' \in \Gamma^*$ ?) Could you design a similar encoding if the target alphabet consisted of only *one* symbol, e.g.  $\Gamma = \{1\}$ ?

**Solution:** Let  $\Sigma$  be some alphabet with  $k$  symbols,  $k > 1$ . The strings of  $\Sigma$  can be coded as strings of  $\Gamma = \{0, 1\}$  in the following manner.

- Set the symbols of  $\Sigma$  to equal integers  $\{1, \dots, k\}$ .
- These numbers (the symbols of  $\Sigma$ ) can be represented with binary numbers of length  $\lceil \log_2 k \rceil$ .
- Every string in  $\Sigma^*$  can therefore be represented as a string of  $\Gamma$  by replacing the symbols of  $\Sigma$  with their binary encoding.

The decoding from  $\Gamma^*$  to  $\Sigma^*$  can be done in a similar fashion by taking strings of length  $\lceil \log_2 k \rceil$  from a string and interpreting them as symbols of  $\Sigma$ .

If the length of a string  $w \in \Sigma^*$  is  $|w| = n$  symbols, the length of its counterpart  $w' \in \Gamma^*$  is  $|w'| = n \cdot \lceil \log_2 k \rceil$ . This is because the number of symbols needed to encode any symbol in  $\Sigma$  is  $\lceil \log_2 k \rceil$ .

For an example, consider the alphabet  $\Sigma = \{a, b, c, d, e, f\}$  and the string  $aacfd$ . As  $|\Sigma| = 6$ ,  $\lceil \log_2 6 \rceil = \lceil 2.58 \rceil = 3$  bits are needed to represent the symbols of  $\Sigma$ . One possible encoding is

$$\begin{array}{ll} a \mapsto 001 & d \mapsto 100 \\ b \mapsto 010 & e \mapsto 101 \\ c \mapsto 011 & f \mapsto 110 \end{array}$$

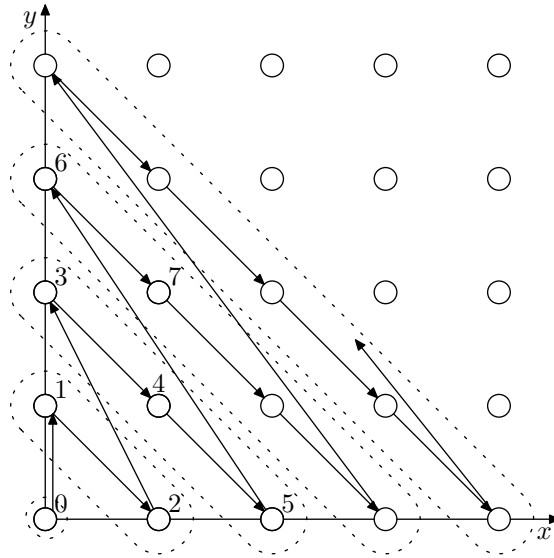
With this encoding, the representation of  $aacfd$  is 001001011110100.

A similar coding scheme cannot be constructed if  $\Gamma = \{1\}$ . A unary presentation of the form  $1 \mapsto 1, 2 \mapsto 11, 3 \mapsto 111, \dots$  can of course be defined, but the code obtained in this way can no longer be decoded unambiguously. For an example, the encodings of 1 1 1, 1 2, 2 1 and 3 are all the string 111.

5. **Problem:** Prove that the Cartesian product  $\mathbb{N} \times \mathbb{N}$  is countably infinite. (*Hint:* Think of the pairs  $(m, n) \in \mathbb{N} \times \mathbb{N}$  as embedded in the Euclidean  $(x, y)$  plane  $\mathbb{R}^2$ . Enumerate the pairs by diagonals parallel to the line  $y = -x$ .) Conclude from this result and the result of Problem 3 that also the set  $\mathbb{Q}$  of rational numbers is countably infinite.

**Solution:** A set  $S$  is countably infinite, if there exists a bijective mapping  $f : \mathbb{N} \rightarrow S$ . By intuition, all members of the set  $S$  can be assigned a unambiguous running number.

The members  $(x, y) \in \mathbb{N} \times \mathbb{N}$  of the set  $\mathbb{N} \times \mathbb{N}$  can be assigned a number as shown in the following figure.



The idea is to arrange all pairs of numbers on diagonals parallel to the line  $y = -x$  and enumerate the lines by member one at a time, starting from the shortest one. Here the enumeration can not be done parallel to the  $x$ -axis; when doing this all indices would be used to enumerate only the  $y$ -axis and no pair  $(x, y), y > 0$  would ever be reached.

The enumerating scheme above can be defined as follows:

$$f(x, y) = x + \sum_{k=1}^{x+y} k = x + \frac{(x+y)(x+y+1)}{2}$$

For an example,  $f(3, 1) = 13$ , that is, the running number of pair  $(3, 1)$  is 13. The function  $f(x, y)$  is a bijection; for every running number there exists a unambiguous pair of numbers. Calculating a coordinate from a given index is relatively difficult, and is discussed in the appendix at the end of these solutions.

The set of rational numbers can be presented as a pair of numbers  $\mathbb{N} \times \mathbb{N}$  by  $(x, y) \equiv \frac{x}{y}, y \neq 0$ . This is a proper subset of the countably infinite set  $\mathbb{N} \times \mathbb{N}$ . By Problem 3,  $\mathbb{Q}$  is either finite or countably infinite. If  $\mathbb{Q}$  was finite, there should exist some rational number  $\frac{x}{y}, x \in \mathbb{N}, y \in \mathbb{N}, y \neq 0$ , that would have the greatest running number  $n < \infty$  (in the enumeration of  $\mathbb{Q}$ ). This cannot be, because using the figure above one could always find a rational number that would have a running number  $n' > n$ . Hence, we have contradiction with the assumption that  $\mathbb{Q}$  is finite. Therefore  $\mathbb{Q}$  is countably infinite.

6. **Problem:** Let  $S$  be an arbitrary nonempty set.

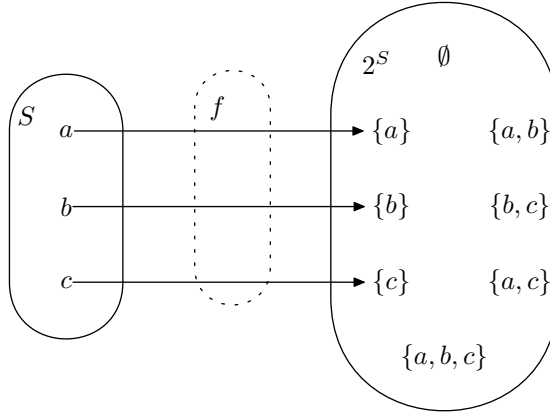
- (a) Give some injective (i.e. one-to-one) function  $f : S \rightarrow \mathcal{P}(S)$ .
- (b) Prove that there cannot exist an injective function  $g : \mathcal{P}(S) \rightarrow S$ . (*Hint:* Assume that such a function  $g$  existed. Consider the set  $R = \{s \in S \mid s \notin g^{-1}(s)\}$ , and denote  $r = g(R)$ . Is it then the case that  $r \in R$ ?)

Observe, as a consequence of item (b), that the power set  $\mathcal{P}(S)$  of any countably infinite set  $S$  is uncountable.

**Solution:** Let  $S \neq \emptyset$  be an arbitrary set.

- (a) Define function  $f : S \rightarrow \mathcal{P}(S)$  so that  $f$  is one-to-one (i.e.  $f$  is an injection). There exists  $\{a\} \in \mathcal{P}(S)$  for all  $a \in S$ . Furthermore, if  $a \neq b$ ,  $\{a\} \neq \{b\}$  holds, so mapping  $f : S \rightarrow \mathcal{P}(S), f(a) = \{a\}$  is such an injection we were looking for.

Presented below is the mapping  $f$  for  $S = \{a, b, c\}$ :



(b) Assume that there exists an injection  $g : \mathcal{P}(S) \rightarrow S$ . Consider the set  $R = \{s \in S \mid s \notin g^{-1}(s)\}$ , and denote  $r = g(R)$ .

If  $r \in R$ ,  $r \notin g^{-1}(r) = g^{-1}(g(R)) = R$  holds. On the other hand, if  $r \notin R$ ,  $r \in g^{-1}(r) = g^{-1}(g(R)) = R$  holds. Both of these lead to contradiction. Therefore it must be so that  $R = \emptyset$  and for all  $s \in S$  it holds that  $s \in g^{-1}(s)$ .

Now  $\emptyset \in \mathcal{P}(S)$ . Injection  $g$  maps  $\emptyset$  to some member of  $S$ :  $g(\emptyset) = x \in S$ . Now  $x \in g^{-1}(x) = \emptyset$  should hold. This is a contradiction. No member can belong to an empty set. (As  $g$  is an injection, the preimage of  $x$  is unambiguous.)

On the basis of (b) an injection  $g : \mathcal{P}(S) \rightarrow S$  cannot be formed. Moreover, if  $S$  is countably infinite, there exists a bijection  $f : \mathbb{N} \rightarrow S$ . For  $\mathcal{P}(S)$  to be countable, there should exist a bijection  $f' : \mathbb{N} \rightarrow \mathcal{P}(S)$ . Assume that such a bijection  $f'$  exists. Then mapping  $g \circ f'^{-1} : \mathcal{P}(S) \rightarrow S$  a bijection (one-to-one and onto). This is contradictory with the fact that there exists no injection  $\mathcal{P}(S) \rightarrow S$ . Therefore  $\mathcal{P}(S)$  is uncountable.

#### Appendix: Counting coordinate pairs from running numbers in problem 4.

Given a running number  $m$  one wishes to calculate such coordinates  $x$  and  $y$  that

$$x + \frac{(x+y)(x+y+1)}{2} = m . \quad (1)$$

Denote  $z = x + y$ . Then (1) equals to:

$$z - y + \frac{z(z+1)}{2} = m . \quad (2)$$

As  $z - y \geq 0$ ,

$$\frac{z(z+1)}{2} \leq m \quad , \text{ it follows that} \quad (3)$$

$$z \leq \frac{-1 + \sqrt{1+8m}}{2} \quad (4)$$

As  $z \in \mathbb{N}$  ja  $m, x, y \geq 0$ , it can be observed that:

$$z = \left\lfloor \frac{-1 + \sqrt{1+8m}}{2} \right\rfloor \quad (5)$$

Now both  $x$  and  $y$  can be calculated using indexing function  $f$ :

$$\begin{aligned} x &= m - f(0, z) \\ y &= z - x \end{aligned} \quad (6)$$

Here  $f(0, z)$  gives the running number of the first member of the diagonal  $(x, y)$ . For an example, let us calculate the pair that corresponds to the running number  $m = 13$ :

$$\begin{aligned}z &= \left\lfloor \frac{-1 + \sqrt{105}}{2} \right\rfloor = \lfloor 4.62 \rfloor = 4 \\x &= 13 - f(0, 4) = 13 - 10 = 3 \\y &= 4 - 3 = 1\end{aligned}$$

As a result the pair  $(3, 1)$  is obtained.