FY2022 "Combinatorics"
Lecture Note

Koji Nuida
nuida@imi.kyushu-u.ac.jp
(Institute of Mathematics for Industry, Kyushu University)

Last Update: August 5, 2022

Contents

1 Bijective Proofs

We consider to prove the following proposition:

Proposition 1.1. *For any integer $n \geq 1$, we have $n! = 1 + \sum_{k=1}^{n-1} k \cdot k!$.*

First Proof. We use mathematical induction. For $n = 1$, both sides in the statement are equal (to 1) as desired. Let $n \geq 2$, and suppose that the claim holds for the case of $n-1$ in order to prove the claim for n. By the claim for $n-1$ we have

$$(n-1)! = 1 + \sum_{k=1}^{n-2} k \cdot k! \ ,$$

and we also have

$$n! - (n-1)! = n \cdot (n-1)! - (n-1)! = (n-1) \cdot (n-1)! \ .$$

By adding these two equalities, we have

$$n! = 1 + \sum_{k=1}^{n-2} k \cdot k! + (n-1) \cdot (n-1)! = 1 + \sum_{k=1}^{n-1} k \cdot k!$$

which is the claim for the case of n. Hence the claim holds for every n. □

This proof is certainly correct, and has an advantage that it is easier to find the strategy of the proof as it goes on a "standard" way using mathematical induction. On the other hand, one may think that this proof does not explain any "intuitive" reason of why this proposition holds.

Next, we consider the following alternative proof for the proposition.

Second Proof. Let S_n denote the symmetric group on n letters. Let id denote the identity permutation (i.e., that fixes every element), and for each $\sigma \in S_n \setminus \{\mathsf{id}\}$, let $k(\sigma)$ denote the maximum integer a with $\sigma(a) \neq a$. Here, each $\sigma \neq \mathsf{id}$ moves at least two elements, therefore we have $\sigma(a) \neq a$ for some

$a \geq 2$. Hence the maximum of such a's indeed exists and we have $k(\sigma) \geq 2$. For each integer k with $1 \leq k \leq n-1$, let

$$X_k := \{\sigma \in S_n \setminus \{\mathsf{id}\} \mid k(\sigma) = k+1\} \ .$$

Then by the argument above, S_n is the disjoint union of subsets $\{\mathsf{id}\}$ and X_k's. Now for any $\sigma \in X_k$, σ fixes every integer larger than or equal to $k+2$. Therefore this σ can be regarded as an element of S_{k+1}, while we have $\sigma(k+1) \neq k+1$ and hence σ is not an element of S_k. By this argument, we have $X_k = S_{k+1} \setminus S_k$ and

$$|X_k| = |S_{k+1}| - |S_k| = (k+1)! - k! = k \cdot k! \ .$$

As S_n is the disjoint union of $\{\mathsf{id}\}$ and X_k's as above, counting the elements yields $|S_n| = 1 + \sum_{k=1}^{n-1} k \cdot k!$, while we have $|S_n| = n!$. Therefore we have $n! = |S_n| = 1 + \sum_{k=1}^{n-1} k \cdot k!$, which implies the claim. □

The key point of the latter proof is the observation that both terms $n!$ and $1 + \sum_{k=1}^{n-1} k \cdot k!$ in the claim represent the number of the elements of the same set, just counted in a different way (hence they are certainly equal). In this proof, the quantities in the given equality are equipped with the "meaning" as the number of elements of a concrete set; one may feel that such a "meaning" helps our intuitive understanding of why this proposition holds. In general, such a technique, of first representing each quantity in a given equality as the number of elements of some concrete set and then showing that the two sets have the same number of elements, is called *bijective proof*, *combinatorial proof*, *counting argument*, etc.

We consider another example of bijective proofs.

Proposition 1.2. *For integers $n \geq k \geq 1$, we have $\binom{n}{k} = \frac{n}{k} \cdot \binom{n-1}{k-1}$, where $\binom{n}{k}$ denotes the binomial coefficient.*

Assuming the explicit expression $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ the proof is obvious, but here we give a proof based on the original definition that $\binom{n}{k}$ is the number of choices of k elements from n elements.

Proof. The claim is equivalent to the equality $k \cdot \binom{n}{k} = n \cdot \binom{n-1}{k-1}$, which we prove from now. Let $[n] := \{1, 2, \ldots, n\}$ and

$$X := \{(I, a) \mid a \in I \subseteq [n]\,,\, |I| = k\}\ .$$

We count the elements of this set X in two ways.

- If we count the elements of X in a way of first choosing I and then choosing a from I, then there are $\binom{n}{k}$ choices of I (by the definition of the binomial coefficient) and k choices of a, hence $|X| = k \cdot \binom{n}{k}$ in total.
- If we count the elements of X in a way of first choosing a and then choosing I involving a, then there are n choices of a and $\binom{n-1}{k-1}$ choices of I (as we have to choose the remaining $k-1$ elements from $n-1$ elements other than a), hence $|X| = n \cdot \binom{n-1}{k-1}$ in total.

They are the left-hand side and the right-hand side in the claim, respectively. Hence the claim holds. □

From now on, we explain a bijective proof for the following theorem called *Euler's pentagonal number theorem*.

Theorem 1.1. *We have* $\displaystyle\prod_{n\geq 1}(1-t^n) = \sum_{k=-\infty}^{\infty}(-1)^k t^{k(3k-1)/2}$.

Formally, both sides of the formula above are regarded as "formal power series (in variable t)"; but here we do not introduce the definition of formal

power series, and we just interpret the claim above as "if we naively expand both sides in a form of power series in t, then for each N, the coefficients of t^N in both sides are equal". For example, when focusing on the coefficient of t^6 in the left-hand side, it is the sum of

- the term $-t^6$ from the product $(1-t^6)$;
- the term $(-t^5)(-t^1) = t^6$ from the product $(1-t^5)(1-t^1)$;
- the term $(-t^4)(-t^2) = t^6$ from the product $(1-t^4)(1-t^2)$; and
- the term $(-t^3)(-t^2)(-t^1) = -t^6$ from the product $(1-t^3)(1-t^2)(1-t^1)$,

resulting in the coefficient being 0. Here we note that in the right-hand side of the claim, when k is a negative value, say $k = -K$, the exponent becomes $(-K)(-3K-1)/2 = K(3K+1)/2$; therefore we can rewrite the right-hand side as

$$1 + \sum_{k=1}^{\infty} (-1)^k t^{k(3k-1)/2} + \sum_{K=1}^{\infty} (-1)^K t^{K(3K+1)/2} \ .$$

We also note that for any positive integers k and K we have $k(3k-1)/2 \neq K(3K+1)/2$. Indeed, if $k \leq K$ then we have $k(3k-1)/2 < K(3K+1)/2$ obviously, while if $k > K$, i.e., $k \geq K+1$ then we have

$$\frac{k(3k-1)}{2} \geq \frac{(K+1)(3K+2)}{2} > \frac{K(3K+1)}{2} \ .$$

This implies that the coefficient of each term when expanding the right-hand side is 0, 1, or -1. In particular, the coefficient of the constant term, i.e., the coefficient of t^0, in the right-hand side is 1, while it is also 1 in the left-hand side as well; therefore it suffices to focus only on the terms with positive exponents.

In order to investigate the expansion of the left-hand side of the claim, we use the following notion.

Definition 1.1. A *partition* means a finite (possibly empty) sequence of positive integers in weakly decreasing order:

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell) \quad (\lambda_i\text{'s are integers with } \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_\ell > 0).$$

More precisely, when such a sequence λ satisfies $|\lambda| := \sum_{i=1}^{\ell} \lambda_i = n$, we say that λ is a partition of integer n and write $\lambda \vdash n$. Here ℓ is called the *length* of the partition λ, denoted by $\ell(\lambda)$. (We express the empty sequence by $\emptyset$; then $\emptyset$ is regarded as the partition consisting of 0 components, with $\emptyset \vdash 0$ and $\ell(\emptyset) = 0$.)

Under the definition, the terms in the left-hand side contributing to the coefficient of t^N are the terms $(-t^{\lambda_1}) \cdots (-t^{\lambda_\ell}) = (-1)^\ell t^N$ coming from the products $(1-t^{\lambda_1}) \cdots (1-t^{\lambda_\ell})$ corresponding to partitions $\lambda = (\lambda_1, \ldots, \lambda_\ell) \vdash N$ with distinct components. Therefore, by expressing by $e(N)$ (respectively, $o(N)$) the number of partitions of N with even (respectively, odd) lengths and distinct components, the coefficient of t^N in the left-hand side becomes $e(N) - o(N)$.

The following notion is useful for visualizing the partitions of integers.

Definition 1.2. Let $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ be a partition. We define the *Young diagram* of λ to be the collection of square boxes arranged in a way that the first row (from the top) has λ_1 boxes, the second row has λ_2 boxes, ..., and the ℓ-th row has λ_ℓ boxes (see Figure 1).

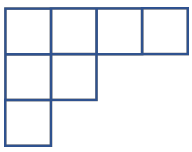


Figure 1: Young diagram of partition $\lambda = (4, 2, 1)$

We note that by the definition of Young diagrams, the number of boxes at each row is weakly decreasing from the top to the bottom. From now

on, let $\mathcal{Y}$ denote the set of non-empty Young diagrams (i.e., those with at least one box) having rows with distinct lengths (i.e., the numbers of boxes). We write the set of Young diagrams $Y \in \mathcal{Y}$ with N boxes as $\mathcal{Y}_N$, and write the set of such Young diagrams with even (respectively, odd) rows as $\mathcal{Y}_{N,e}$ (respectively, $\mathcal{Y}_{N,o}$). Then we have $e(N) - o(N) = |\mathcal{Y}_{N,e}| - |\mathcal{Y}_{N,o}|$. Intuitively, the equality in the claim is interpreted as that the numbers of elements of $\mathcal{Y}_{N,e}$ and $\mathcal{Y}_{N,o}$ are "approximately" equal. In order to analyze the "error" in the approximation, we define the following Young diagrams. For an integer $k \geq 1$, let $\widehat{Y}_k$ denote the Young diagram with k rows for which the first row has $2k-1$ boxes and the lengths of rows are decremented by 1, hence the last, k-th row has k boxes. Similarly, for an integer $K \geq 1$, let $\widetilde{Y}_K$ denote the Young diagram with K rows for which the first row has $2K$ boxes and the lengths of rows are decremented by 1, hence the last, K-th row has $K+1$ boxes. Let

$$\mathcal{E} := \{\widehat{Y}_k \mid k \geq 1\} \cup \{\widetilde{Y}_K \mid K \geq 1\} \ .$$

We note that the numbers of boxes of $\widehat{Y}_k$ and $\widetilde{Y}_K$ are $((2k-1)+k) \cdot k/2 = k(3k-1)/2$ and $(2K + (K+1)) \cdot K/2 = K(3K+1)/2$, respectively. Now when $k \geq 1$ is odd, by setting $N = k(3k-1)/2$, the coefficient of t^N in the right-hand side of the claim becomes -1, therefore we have to show that $|\mathcal{Y}_{N,e}| - |\mathcal{Y}_{N,o}| = -1$ or equivalently $|\mathcal{Y}_{N,e}| = |\mathcal{Y}_{N,o}| - 1$. For the proof, it suffices to show that there exists a bijection between the set $\mathcal{Y}_{N,o}$ excluded the "exception" $\widehat{Y}_k$ and the set $\mathcal{Y}_{N,e}$ (hence these sets have the same number of elements). On the other hand, when $k \geq 1$ is even, by setting $N = k(3k-1)/2$, the coefficient of t^N in the right-hand side of the claim becomes 1, therefore we have to show that $|\mathcal{Y}_{N,e}| - |\mathcal{Y}_{N,o}| = 1$ or equivalently $|\mathcal{Y}_{N,e}| - 1 = |\mathcal{Y}_{N,o}|$. For the proof, it suffices to show that there exists a bijection between the set $\mathcal{Y}_{N,e}$ excluded the "exception" $\widehat{Y}_k$ and the set $\mathcal{Y}_{N,e}$. Similarly, for the comparison of the coefficients of t^N with $N = K(3K+1)/2$, it suffices to compare the sets $\mathcal{Y}_{N,e}$ and $\mathcal{Y}_{N,o}$ where $\widetilde{Y}_K$ is treated as the "exception". For the other N's, we need not to consider such exceptions, and it suffices to show

the existence of a bijection simply between $\mathcal{Y}_{N,e}$ and $\mathcal{Y}_{N,o}$. Summarizing, our task is to show the following: for each $N \geq 1$, there exists a bijection between $\mathcal{Y}_{N,e} \setminus \mathcal{E}$ and $\mathcal{Y}_{N,o} \setminus \mathcal{E}$.

For each $Y \in \mathcal{Y}$, we write the length of the last row of Y as $a = a(Y)$, and write as $b = b(Y)$ the unique k satisfying that the lengths of Y's first to k-th rows are decreased by 1 and the length of Y's $(k+1)$-th row is decreased by at least 2 from the k-th row. Here we note that we set $b(Y) = 1$ when the lengths of the first and the second rows differ by at least 2, and we set $b(Y)$ to be the number of rows of Y when the lengths of all the rows are decreased by 1 (see Figure 2). For example, for the Young diagram Y of $\lambda = (7, 6, 4, 3)$, we have $a(Y) = 3$ and $b(Y) = 2$. Now we define $\varphi(Y)$ to be the element of $\mathcal{Y}$ obtained by the following procedure:

- When $a \leq b$, move the boxes at the last row of Y to (the right of) the first a rows, one box per each row (see the left part of Figure 3).
- When $a > b$, move the rightmost boxes of the first b rows of Y to the bottom of the diagram, i.e., as the row next to the last row of Y (see the right part of Figure 3).

Here, when $a \leq b$, if Y has precisely b rows and $a = b$ (i.e., $Y = \widehat{Y}_b$) then the operation is not well-defined (as the rows with deletion and with insertion overlap); we omit such cases from the operation. Similarly, when $a > b$, if Y has precisely b rows and $a = b + 1$ (i.e., $Y = \widetilde{Y}_b$) then the operation is not well-defined (as the b-th and the $(b + 1)$-th rows after the operation will both have length b); we omit such cases from the operation. That is, the operation φ is in fact defined for elements of $\mathcal{Y} \setminus \mathcal{E}$. Now we have the following properties.

Lemma 1.1. *For any $Y \in \mathcal{Y} \setminus \mathcal{E}$, we have $\varphi(Y) \in \mathcal{Y} \setminus \mathcal{E}$.*

Proof. First we show that $\varphi(Y) \in \mathcal{Y}$. When $a(Y) \leq b(Y)$, the property follows from the fact that the operation φ does not decrease the differences

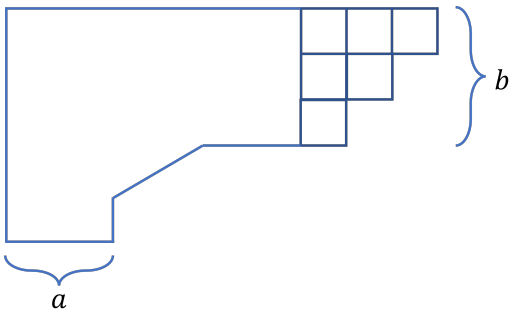


Figure 2: Sketch of the definitions of $a = a(Y)$ and $b = b(Y)$

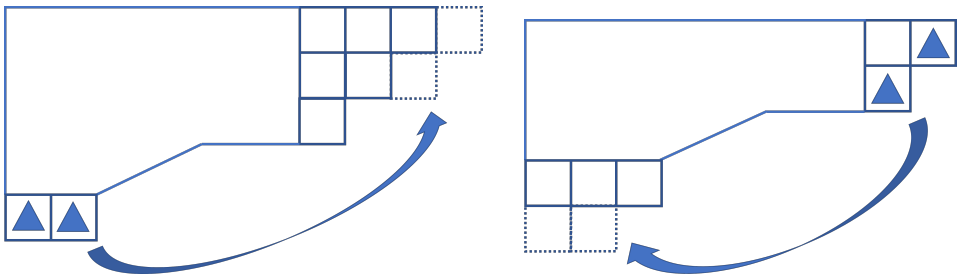


Figure 3: Sketch of the definition of operation φ (the left part is for $a \leq b$, and the right part is for $a > b$)

of lengths between the rows. When $a(Y) > b(Y)$, the difference of lengths between two consecutive rows is changed by φ only at the $b(Y)$-th and the $(b(Y)+1)$-th rows and at the newly introduced row and the original last row. If Y has precisely $b(Y)$ rows (note that now $a(Y) \geq b(Y) + 2$ as $Y \notin \mathcal{E}$), then the property holds as the length of the newly introduced row is $b(Y) < a(Y) - 1$. In the other case, for the former pair of rows, the lengths are still different after the operation as the difference of the lengths in the original Y is at least 2. For the latter pair of rows, the lengths of the newly introduced row being $b(Y)$ and of the original last row being $a(Y)$ are indeed difference by the condition above. Hence we have $\varphi(Y) \in \mathcal{Y}$ in any case.

From now, we show that $\varphi(Y) \notin \mathcal{E}$. Assume for the contrary that the lengths of all the rows of $\varphi(Y)$ are decremented by 1. If $a(Y) \leq b(Y)$, then the phenomenon above happens only when $b(Y) = a(Y)$. In this case, the

condition $Y \notin \mathcal{E}$ implies that Y has at least $b(Y)+1$ rows. Now the length of Y's $b(Y)$-th row is at least $a(Y)+2 = b(Y)+2$, therefore the length of $\varphi(Y)$'s $b(Y)$-th row is at least $b(Y)+3$. However, such an element of $\mathcal{E}$ does not exist, a contradiction. On the other hand, if $a(Y) > b(Y)$, then $\varphi(Y)$ has at least $b(Y)+1$ rows and its last row has length $b(Y)$. However, such an element of $\mathcal{E}$ does not exist, a contradiction. Hence we have a contradiction in any case, therefore we have $\varphi(Y) \notin \mathcal{E}$, as desired. Hence the claim holds. □

Lemma 1.2. *For any $Y \in \mathcal{Y} \setminus \mathcal{E}$, we have $\varphi(\varphi(Y)) = Y$.*

Proof. First we consider the case $a(Y) > b(Y)$. By the definition of φ, we have $a(\varphi(Y)) = b(Y)$ and $b(\varphi(Y)) \geq b(Y)$, therefore $a(\varphi(Y)) \leq b(\varphi(Y))$. Now the operation φ for $\varphi(Y)$ moves the $b(Y)$ boxes in the last row, which were moved by the first operation φ, back to the first $b(Y)$ rows; hence we have $\varphi(\varphi(Y))$.

Secondly, we consider the other case $a(Y) \leq b(Y)$. When Y has at least $b(Y)+1$ rows, let c be the length of the second last row of Y. Then by the definition of φ, we have $a(\varphi(Y)) \geq c > a(Y)$ and $b(\varphi(Y)) = a(Y)$, therefore $a(\varphi(Y)) > b(\varphi(Y))$. Now the operation φ for $\varphi(Y)$ moves the rightmost boxes in the first $a(Y)$ rows, which were moved by the first operation φ, back to the bottom; hence we have $\varphi(\varphi(Y))$.

On the other hand, when Y has precisely $b(Y)$ rows, the condition $Y \notin \mathcal{E}$ implies that $a(Y) < b(Y)$. Then by the definition of φ, we have $a(\varphi(Y)) \geq a(Y)+1$ and $b(\varphi(Y)) = a(Y)$, therefore $a(\varphi(Y)) > b(\varphi(Y))$. Now the operation φ for $\varphi(Y)$ moves the rightmost boxes in the first $a(Y)$ rows, which were moved by the first operation φ, back to the bottom; hence we have $\varphi(\varphi(Y))$. Therefore the claim holds in any case. □

By Lemma 1.2, φ is the inverse map of itself, therefore it is a bijection from $\mathcal{Y} \setminus \mathcal{E}$ to the same set. Moreover, by the definition of φ, it does not change the total number of boxes and increases the number of rows by 1, therefore we have $\varphi(\mathcal{Y}_{N,e} \setminus \mathcal{E}) \subseteq \mathcal{Y}_{N,o} \setminus \mathcal{E}$ and $\varphi(\mathcal{Y}_{N,o} \setminus \mathcal{E}) \subseteq \mathcal{Y}_{N,e} \setminus \mathcal{E}$. Now the

surjectivity of φ implies that $\varphi(\mathcal{Y}_{N,e} \setminus \mathcal{E}) = \mathcal{Y}_{N,o} \setminus \mathcal{E}$ and $\varphi(\mathcal{Y}_{N,o} \setminus \mathcal{E}) = \mathcal{Y}_{N,e} \setminus \mathcal{E}$. Summarizing, it follows for each N that φ is a bijection between $\mathcal{Y}_{N,e} \setminus \mathcal{E}$ and $\mathcal{Y}_{N,o} \setminus \mathcal{E}$, as desired. This completes the proof of Theorem 1.1.

Exercises

Problem 1. For integers $n \geq k \geq 0$, give a bijective proof of the expression of the binomial coefficient $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$.

(Hint: By multiplying the denominator, the target equality becomes $k!(n-k)! \cdot \binom{n}{k} = n!$. Its right-hand side means the number of elements of S_n; in what way can we obtain the left-hand side by counting the elements of S_n?)

Problem 2. For integer $n \geq 1$, give a bijective proof of the equality $\displaystyle\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$.

(Hint: We can rewrite the claim as $\displaystyle\sum_{k;\,\text{even}} \binom{n}{k} = \sum_{k;\,\text{odd}} \binom{n}{k}$.)

2 Generating Functions

As an example, we consider the sequence $(b_n)_n$ defined by the following recurrence relation:

$$b_0 = b_1 = 1\,,\ b_n = b_{n-1} + b_{n-2} \quad (n \geq 2).$$

It is well-known that this defines the Fibonacci numbers, for which the general term can be explicitly computed by an elementary method. Here, as an alternative approach, we are trying to compute the general term by using a tool called *generating functions*. For the purpose, we introduce the generating function for the sequence $(b_n)_n$ as follows:

$$B(t) := \sum_{n \geq 0} b_n t^n \ .$$

Formally, this is a kind of "formal power series"; here we postpone the detailed explanation and just regard it as "a formal sum similar to polynomials". In order to derive a "functional equation" for this generating function, we (heuristically) consider $\widehat{B}(t) := B(t) - tB(t) - t^2 B(t)$. For $n \geq 2$, computation of the coefficient of t^n in $\widehat{B}(t)$ yields

$$\begin{aligned}
&[\text{coeff. of } t^n \text{ in } B(t)] - [\text{coeff. of } t^n \text{ in } tB(t)] - [\text{coeff. of } t^n \text{ in } t^2 B(t)] \\
&= [\text{coeff. of } t^n \text{ in } B(t)] - [\text{coeff. of } t^{n-1} \text{ in } B(t)] - [\text{coeff. of } t^{n-2} \text{ in } B(t)] \\
&= b_n - b_{n-1} - b_{n-2} = 0
\end{aligned}$$

(where the last equality follows from the recurrence relation). Therefore, $\widehat{B}(t)$ is a polynomial in t of degree at most 1. By computing the coefficients of t^1 and t^0 similarly, we have

$$\begin{aligned}
&[\text{coeff. of } t^1 \text{ in } \widehat{B}(t)] \\
&= [\text{coeff. of } t^1 \text{ in } B(t)] - [\text{coeff. of } t^1 \text{ in } tB(t)] - [\text{coeff. of } t^1 \text{ in } t^2 B(t)] \\
&= [\text{coeff. of } t^1 \text{ in } B(t)] - [\text{coeff. of } t^0 \text{ in } B(t)] - 0 \\
&= b_1 - b_0 = 0 \ ,
\end{aligned}$$

$$\begin{aligned}
&[\text{coeff. of } t^0 \text{ in } \widehat{B}(t)] \\
&= [\text{coeff. of } t^0 \text{ in } B(t)] - [\text{coeff. of } t^0 \text{ in } tB(t)] - [\text{coeff. of } t^0 \text{ in } t^2B(t)] \\
&= [\text{coeff. of } t^0 \text{ in } B(t)] - 0 - 0 \\
&= b_0 = 1 \ .
\end{aligned}$$

Summarizing, we have $\widehat{B}(t) = B(t) - tB(t) - t^2B(t) = 1$, therefore $B(t)$ can be written as

$$B(t) = \frac{1}{1-t-t^2} \ .$$

(Here we postpone the explanation of the meaning of the rational expression in the right-hand side as "formal power series".)

In order to determine the coefficient of each monomial in $B(t)$ (i.e., the general term of $(b_n)_n$) from the expression above, we perform a partial fraction decomposition by using the relation $1-t-t^2 = \left(1 - \frac{1+\sqrt{5}}{2}t\right)\left(1 - \frac{1-\sqrt{5}}{2}t\right)$:

$$B(t) = \frac{1}{1-t-t^2} = \frac{1}{\sqrt{5}t}\left(\frac{1}{1-\frac{1+\sqrt{5}}{2}t} - \frac{1}{1-\frac{1-\sqrt{5}}{2}t}\right) \ .$$

By naively applying the formula for the sums of geometric progressions, the right-hand side becomes

$$\frac{1}{\sqrt{5}t}\left(\sum_{k\geq 0}\left(\frac{1+\sqrt{5}}{2}t\right)^k - \sum_{k\geq 0}\left(\frac{1-\sqrt{5}}{2}t\right)^k\right) \ .$$

As the coefficient of t^n is b_n, by dividing the coefficient of t^{n+1} in the parentheses by $\sqrt{5}$, we have

$$b_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right) \ .$$

We have thus determined the general term of $(b_n)_n$.

Besides the example above that can be already solved elementarily, we give another example that is more complicated. For integer $n \geq 0$, we define

the n-th *Catalan number* C_n to be the number of ways of moving from the point $(0,0)$ to the point $(2n,0)$ on the coordinate plane subject to the following rule:

- Each step is either "1 in the direction of x-axis, 1 in the direction of y-axis" or "1 in the direction of x-axis, -1 in the direction of y-axis".
- The y-coordinate should be kept non-negative during the move.

Any path satisfying these conditions is called a *Dyck path*. For example, when $n = 3$, by expressing the step $(1,1)$ and $(1,-1)$ as $\nearrow$ and $\searrow$, respectively, there are in total the following 5 Dyck paths from $(0,0)$ to $(6,0)$:

$\nearrow\nearrow\nearrow\searrow\searrow\searrow$, $\nearrow\nearrow\searrow\nearrow\searrow\searrow$, $\nearrow\nearrow\searrow\searrow\nearrow\searrow$, $\nearrow\searrow\nearrow\nearrow\searrow\searrow$, $\nearrow\searrow\nearrow\searrow\nearrow\searrow$,

therefore $C_3 = 5$. Note that $C_0 = 1$ by regarding the case of $n = 0$ as having the empty path as the unique Dyck path. We try to determine the general term of Catalan numbers by using the generating function $C(t) := \sum_{n\geq 0} C_n t^n$.

We first give a recurrence relation for Catalan numbers. For the purpose, for $n \geq 1$, we focus on the point at which a given Dyck path firstly intersects with the x-axis (i.e., the y-coordinate becomes 0) except for the starting point $(0,0)$. By considering the possible patterns of the change of y-coordinate, we can observe that such a point must be of the form $(2k,0)$ (where k is an integer and $1 \leq k \leq n$). Now for such a Dyck path, the first step is $\nearrow$, the $2k$-th step is $\searrow$, and the second to the $(2k-1)$-th steps give "a move from $(1,1)$ to $(2k-1,1)$ while keeping the y-coordinate not less than 1". By a translation in the direction of $(-1,-1)$, the path indicated by "..." corresponds to a Dyck path from $(0,0)$ to $(2k-2,0)$. On the other hand, the remaining part of the path from $(2k,0)$ to $(2n,0)$ corresponds to a Dyck path from $(0,0)$ to $(2n-2k,0)$ by a translation in the direction of $(-2k,0)$. The whole path is determined by the pair of those two paths. As the value of k

varies over the range $1 \leq k \leq n$, we obtain the following recurrence relation:

$$C_0 = 1\,,\ C_n = \sum_{k=1}^{n} C_{k-1}C_{n-k} \quad (n \geq 1).$$

Now we consider $\widehat{C}(t) := C(t) - tC(t)^2$. For $n \geq 1$, we have

$$\begin{aligned}
[\text{coeff. of } t^n \text{ in } \widehat{C}(t)] &= [\text{coeff. of } t^n \text{ in } C(t)] - [\text{coeff. of } t^n \text{ in } tC(t)^2] \\
&= C_n - [\text{coeff. of } t^{n-1} \text{ in } C(t)^2] \\
&= C_n - \sum_{k=0}^{n-1} [\text{coeff. of } t^k \text{ in } C(t)] \cdot [\text{coeff. of } t^{n-1-k} \text{ in } C(t)] \\
&= C_n - \sum_{k=0}^{n-1} C_k C_{n-1-k} \\
&= C_n - \sum_{k=1}^{n} C_{k-1}C_{n-k} = 0
\end{aligned}$$

(where the last equality follows from the recurrence relation above). This implies that $\widehat{C}(t)$ is a constant, while the coefficient of t^0 is $C_0 - 0 = 1$, therefore we have $\widehat{C}(t) = C(t) - tC(t)^2 = 1$ and

$$tC(t)^2 - C(t) + 1 = 0\ .$$

In order to solve the quadratic equation in $C(t)$ above, by multiplying both sides by $4t$ we have

$$4t^2C(t)^2 - 4tC(t) + 4t = 0\ ,$$

which can be transformed as

$$(2tC(t) - 1)^2 = 1 - 4t\ .$$

To obtain a square root of $1 - 4t$ in the right-hand side, we prepare the following lemma.

Lemma 2.1. *For any real number λ, define*

$$F_\lambda(t) := \sum_{n \geq 0} \frac{\lambda(\lambda-1)\cdots(\lambda-n+1)}{n!} t^n \ .$$

Then we have $F_\lambda(t)F_\mu(t) = F_{\lambda+\mu}(t)$.

Proof. Let $P_n(\lambda,\mu)$ and $Q_n(\lambda,\mu)$ denote the coefficients of t^n in the left-hand and the right-hand sides of the claim, respectively. They are both elements of the polynomial ring $\mathbb{R}[\lambda,\mu]$. Let $R_n(\lambda,\mu) := P_n(\lambda,\mu) - Q_n(\lambda,\mu)$. Now if we fix any integer $\mu \geq 0$, then $R_n(\lambda,\mu)$ is regarded as an element of $\mathbb{R}[\lambda]$. In this setting, the terms in $F_\mu(t)$ of degree $\mu+1$ or higher all have coefficients 0, therefore we have

$$F_\mu(t) = \sum_{n=0}^{\mu} \binom{\mu}{n} t^n = (1+t)^\mu \ .$$

If moreover λ is also a non-negative integer, then we have $F_\lambda(t) = (1+t)^\lambda$ and $F_{\lambda+\mu}(t) = (1+t)^{\lambda+\mu}$ similarly, therefore $F_\lambda(t)F_\mu(t) = F_{\lambda+\mu}(t)$, hence $P_n(\lambda,\mu) = Q_n(\lambda,\mu)$ or equivalently $R_n(\lambda,\mu) = 0$. This implies that when fixing any integer $\mu \geq 0$, the polynomial $R_n(\lambda,\mu)$ in λ has value 0 at every non-negative integer point, implying that it is the zero polynomial. Accordingly, we have $R_n(\lambda,\mu) = 0$ for any $\lambda \in \mathbb{R}$ and any non-negative integer μ. Then by fixing any real number λ, the polynomial $R_n(\lambda,\mu)$ in μ has value 0 at every non-negative integer point, implying that it is the zero polynomial. Accordingly, we have $R_n(\lambda,\mu) = 0$ for any $\lambda,\mu \in \mathbb{R}$. This means that $P_n(\lambda,\mu)$ and $Q_n(\lambda,\mu)$ always coincide, therefore the definitions of P_n and Q_n imply the desired equality $F_\lambda(t)F_\mu(t) = F_{\lambda+\mu}(t)$. □

By Lemma 2.1 we have $F_{1/2}(t)^2 = F_1(t) = 1+t$; by substituting $-4t$ into t we have $F_{1/2}(-4t)^2 = 1-4t$. This implies that

$$2tC(t) - 1 = \pm F_{1/2}(-4t) \tag{1}$$

(note that the set of all formal power series forms an integral domain as explained above, therefore any element has at most two square roots). Moreover, we have

$$\begin{aligned} F_{1/2}(-4t) &= \sum_{n\geq 0} \frac{1/2 \cdot (1/2-1) \cdots (1/2-n+1)}{n!}(-4t)^n \\ &= 1 + \frac{1}{2}\cdot(-4t) + \sum_{n\geq 2} \frac{1\cdot(-1)\cdot(-3)\cdots(-(2n-3))}{2^n \cdot n!}(-4)^n t^n \ , \end{aligned}$$

while for $n \geq 2$, we have

$$\begin{aligned} \frac{1\cdot(-1)\cdot(-3)\cdots(-(2n-3))}{2^n \cdot n!}(-4)^n &= \frac{(-1)^{n-1}\cdot 1\cdot 1\cdot 3\cdots(2n-3)}{n!}(-2)^n \\ &= -\frac{(2n-2)!}{2\cdot 4\cdots(2n-2)\cdot n!}2^n \\ &= -\frac{(2n-2)!}{2^{n-1}\cdot 1\cdot 2\cdots(n-1)\cdot n!}2^n \\ &= -\frac{2\cdot(2n-2)!}{(n-1)!n!} \ , \end{aligned}$$

therefore we have

$$F_{1/2}(-4t) = 1 - 2t + \sum_{n\geq 2}(-2)\frac{(2n-2)!}{(n-1)!n!}t^n \ .$$

Now comparison of the constant terms in both sides of Eq.(1) determines the sign $\pm$, which yields

$$2tC(t) - 1 = -F_{1/2}(-4t)$$

and

$$2tC(t) = 1 - F_{1/2}(-4t) = 2t + \sum_{n\geq 2} 2\frac{(2n-2)!}{(n-1)!n!}t^n = \sum_{n\geq 1} 2\frac{(2n-2)!}{(n-1)!n!}t^n \ .$$

Dividing both sides by $2t$ yields the generating function $C(t)$ given by

$$C(t) = \sum_{n\geq 1}\frac{(2n-2)!}{(n-1)!n!}t^{n-1} = \sum_{n\geq 0}\frac{(2n)!}{n!(n+1)!}t^n \ .$$

Hence the general term of Catalan numbers is determined as

$$C_n = \frac{(2n)!}{n!(n+1)!} = \frac{1}{n+1}\binom{2n}{n} \ .$$

We note that there are a significantly large number of other equivalent combinatorial characterizations for Catalan numbers; we refer for them to Chapter 6, Exercise 6.19 of Stanlay's book [6] (see also an exercise below).

From now on, we consider the question "what is a formal power series?" postponed from the argument above. A possible approach, similar to the formal definition of polynomial rings, is focusing on the maps with domains being the set $\{0, 1, 2, \dots\}$ of exponents, defining addition and multiplication for the set of all such maps appropriately, and so on. Here we adopt a different approach. Roughly speaking, as an analogy of analytically convergent power series being the limits of polynomial sequences, we want to deal with (not necessarily convergent) formal power series as "the limits of polynomials". For the purpose, we prepare some definitions and properties.

Definition 2.1. Let K be a field. We say that a map $\nu\colon K \to \mathbb{R}_{\geq 0}$ is a *non-Archimedean valuation* on K if the following conditions hold:

1. For any $x \in K$, the conditions $\nu(x) = 0$ and $x = 0$ are equivalent.
2. For any $x, y \in K$, we have $\nu(xy) = \nu(x)\nu(y)$.
3. For any $x, y \in K$, we have $\nu(x+y) \leq \max\{\nu(x), \nu(y)\}$.

We simply call a non-Archimedean valuation a *valuation* in the following argument. The following general result is well-known (we omit the proof here).

Proposition 2.1. *Let $\nu\colon K \to \mathbb{R}_{\geq 0}$ be a valuation on K, and define $d\colon K \times K \to \mathbb{R}$ by $d(x, y) := \nu(x - y)$. Then d is a metric on K. Let $\widehat{K}$ be the completion of K with respect to this metric. Then $\widehat{K}$ is naturally an extension field of K, and ν is extendible to a valuation $\widehat{\nu}$ on $\widehat{K}$. Moreover, the set $\{\widehat{x} \in \widehat{K} \mid \widehat{\nu}(\widehat{x}) \leq 1\}$ forms a complete subring of $\widehat{K}$.*

Let K be a field. We are going to define a valuation ν on the univariate rational function field $K(t)$ over K. First, we define $\nu(0) := 0$, and for any polynomial $f(t) = a_d t^d + \cdots + a_{d+\ell} t^{d+\ell}$ ($a_d \neq 0$), define $\nu(f) := e^{-d}$. Based on this, for $h(t) = f(t)/g(t)$ ($f(t), g(t) \in K[t]$, $g \neq 0$) we define $\nu(h) := \nu(f)/\nu(g)$. Then the value of ν is well-defined regardless of the expression $f(t)/g(t)$ of $h(t)$.

Proposition 2.2. *The ν above is a valuation on $K(t)$.*

Proof. The fact that ν is a map to $\mathbb{R}_{\geq 0}$ and Condition 1 in the definition are deduced from the argument above and the property $e^{-d} > 0$. Condition 2 is deduced by observing that if $f(t) = a_d t^d + \cdots + a_{d+\ell} t^{d+\ell}$ ($a_d \neq 0$) and $g(t) = b_{d'} t^{d'} + \cdots + b_{d'+\ell'} t^{d'+\ell'}$ ($b_{d'} \neq 0$) then we have $f(t)g(t) = a_d b_{d'} t^{d+d'} + \cdots$, $a_d b_{d'} \neq 0$, and $e^{-(d+d')} = e^{-d} e^{-d'}$. For Condition 3, first, for $f(t) = a_d t^d + \cdots + a_{d+\ell} t^{d+\ell}$ ($a_d \neq 0$) and $g(t) = b_{d'} t^{d'} + \cdots + b_{d'+\ell'} t^{d'+\ell'}$ ($b_{d'} \neq 0$), the coefficients in $f(t)+g(t)$ of all the terms with degree less than $\min\{d, d'\}$ are 0. Therefore we have $\nu(f+g) \leq e^{-\min\{d,d'\}} = \max\{e^{-d}, e^{-d'}\} = \max\{\nu(f), \nu(g)\}$. This implies that the right-hand side of

$$\nu\left(\frac{f_1}{g_1} + \frac{f_2}{g_2}\right) = \nu\left(\frac{f_1 g_2 + f_2 g_1}{g_1 g_2}\right) = \frac{\nu(f_1 g_2 + f_2 g_1)}{\nu(g_1)\nu(g_2)}$$

becomes

$$\begin{aligned} \leq \frac{\max\{\nu(f_1 g_2), \nu(f_2 g_1)\}}{\nu(g_1)\nu(g_2)} &= \max\left\{\frac{\nu(f_1 g_2)}{\nu(g_1)\nu(g_2)}, \frac{\nu(f_2 g_1)}{\nu(g_1)\nu(g_2)}\right\} \\ &= \max\left\{\frac{\nu(f_1)}{\nu(g_1)}, \frac{\nu(f_2)}{\nu(g_2)}\right\} = \max\{\nu(f_1/g_1), \nu(f_2/g_2)\} \ , \end{aligned}$$

therefore Condition 3 holds. Hence the claim holds. □

By applying Proposition 2.1 to this valuation ν on $K(t)$, we obtain the valuation $\widehat{\nu}$ on $\widehat{K(t)}$. Now the set

$$k[[t]] := \{F \in \widehat{K(t)} \mid \widehat{\nu}(F) \leq 1\}$$

is a complete subring of the field $\widehat{K(t)}$, hence in particular it is an integral domain. We call the $K[[t]]$ the (univariate) *formal power series ring*, and call its elements *formal power series*.

In order to show that the formal power series are "limits of polynomials", we prepare the following lemma.

Lemma 2.2. *Let X be a metric space, let $(x_n)_n$ be a Cauchy sequence in X, and let $(y_m^{(n)})_m$ (for $n \geq 0$) be sequences in X that uniformly converge to x_n (that is, for any $\varepsilon > 0$, there exists an N satisfying that for any n and any $m \geq N$ we have $d(y_m^{(n)}, x_n) \leq \varepsilon$). Then $(y_n^{(n)})_n$ is a Cauchy sequence equivalent to $(x_n)_n$.*

Proof. We show that the two sequences in the statement are equivalent. Let $\varepsilon > 0$. By using the definition for $(x_n)_n$ being a Cauchy sequence, with parameter $\varepsilon/2 > 0$, it follows that there exists an N_1 satisfying that for any $m, n \geq N_1$ we have $d(x_m, x_n) \leq \varepsilon/2$. On the other hand, by using the definition for $(y_m^{(n)})_m$ uniformly converging to x_n, with parameter $\varepsilon/2 > 0$, it follows that there exists an N_2 satisfying that for any n and any $m \geq N_2$ we have $d(y_m^{(n)}, x_n) \leq \varepsilon/2$; hence by setting $n = m$ we have $d(y_m^{(m)}, x_m) \leq \varepsilon/2$. These arguments imply that for any $m, n \geq \max\{N_1, N_2\}$ we have

$$d(y_m^{(m)}, x_n) \leq d(y_m^{(m)}, x_m) + d(x_m, x_n) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

(where we used the triangle inequality at the first inequality). This means that $(y_n^{(n)})_n$ is equivalent to $(x_n)_n$. Moreover, the property that $(y_n^{(n)})_n$ is a Cauchy sequence follows from the fact that any sequence equivalent to a Cauchy sequence is also a Cauchy sequence. Hence the claim holds. □

We note that the polynomial ring $K[t]$ is a subring of $K(t)$, therefore it is also a subring of $\widehat{K(t)}$.

Theorem 2.1. *$K[t]$ is a dense subring of $K[[t]]$.*

Proof. The property $K[t] \subseteq K[[t]]$ follows from the definition of the valuation ν on $K(t)$ and the fact that $e^{-d} \leq 1$ for any $d \geq 0$.

Take any element of $K[[t]]$ and the corresponding Cauchy sequence $(h_n)_n$ in $K(t)$. We have $\lim_{n\to\infty} \nu(h_n) \leq 1$ by the definition of $K[[t]]$. As 1 is an isolated point of the image of $\nu\colon K(t) \to \mathbb{R}_{\geq 0}$, we have $\nu(h_n) \leq 1$ for any sufficiently large n. We may assume without loss of generality that $\nu(h_n) \leq 1$ for any n, by replacing $(h_n)_n$ with an appropriate equivalent sequence if necessary. Then when expressing h_n as an irreducible fraction of polynomials, the constant term of the denominator is non-zero. Hence we can write $h_n = f_n/(1+tg_n)$ with $f_n, g_n \in K[t]$.

Now let $F_{n,m} := f_n \sum_{k=0}^{m} (-tg_n)^k$. We show that the sequence $(F_{n,m})_m$ uniformly converges to h_n. For any n, m, we have

$$\begin{aligned} F_{n,m} - h_n &= \frac{f_n}{1+tg_n} \left((1+tg_n) \sum_{k=0}^{m} (-tg_n)^k - 1 \right) \\ &= h_n \cdot (1 + (-1)^m (tg_n)^{m+1} - 1) = (-1)^m h_n \cdot (tg_n)^{m+1} \ , \end{aligned}$$

therefore we have

$$\nu(F_{n,m} - h_n) = \nu(h_n)\nu(t)^{m+1}\nu(g_n)^{m+1} \leq 1 \cdot e^{-m-1} \cdot 1^{m+1} = e^{-m-1} \ .$$

The right-hand side is independent of n and converges to 0. Therefore $(F_{n,m})_m$ uniformly converges to h_n.

Now by Lemma 2.2, the element of $K[[t]]$ chosen firstly is also the limit of Cauchy sequence $(F_{n,n})_n$. As each $F_{n,n}$ is an element of $K[t]$, it follows that $K[t]$ is dense in $K[[t]]$. Hence the claim holds. □

When $f_n = a_{n,0} + a_{n,1}t + \cdots + a_{n,d_n}t^{d_n} \in K[t]$, the condition for $(f_n)_n$ being a Cauchy sequence is equivalent to that for any $k \geq 0$, $a_{n,k}$ is constant for sufficiently large n's (where we set $a_{n,k} = 0$ for $k > d_n$). By writing the "constant" as $\widehat{a}_k \in K$ and expressing the element of $K[[t]]$ obtained as the limit of $(f_n)_n$ as $\sum_{n\geq 0} \widehat{a}_n t^n$, we obtain a "formal power series" that looks similarly as in the previous arguments. We note that when $(f_n)_n$ and $(g_n)_n$ are

equivalent Cauchy sequences, the elements $\widehat{a}_k$ obtained as above from those sequences become equal to each other, therefore the expression $\sum_{n\geq 0} \widehat{a}_n t^n$ of a formal power series is uniquely determined. We also note that for any sequence $(a_n)_n$, the sequence $(f_n)_n$ with $f_n := \sum_{k=0}^{n} a_k t^k$ is a Cauchy sequence, therefore the formal power series $\sum_{n\geq 0} a_n t^n$ (i.e., the generating function of $(a_n)_n$) obtained as the limit of $(f_n)_n$ always exists. Moreover, as the formal power series are expressed as limits of polynomials, the addition, subtraction, and multiplication operations for polynomials are extended naturally to formal power series.

By defining formal power series as above, several formal operations for formal power series performed in the previous arguments can be justified. For example:

- For the infinite product $\prod_{n\geq 1}(1-t^n)$ appeared in Euler's pentagonal number theorem, the sequence $(f_n)_n$ with $f_n := \prod_{k=1}^{n}(1-t^k)$ forms a Cauchy sequence with respect to the metric above (note that multiplying $1-t^n$ to a polynomial does not change the coefficients of the terms of degrees less than n), therefore the formal power series $\prod_{n\geq 1}(1-t^n)$ is well-defined as the limit of $(f_n)_n$.

- For a polynomial of the form $1-tf$ ($f \in K[t]$), the sequence $(g_n)_n$ defined by $g_n := \sum_{k=0}^{n}(tf)^k$ forms a Cauchy sequence. Let $g \in K[[t]]$ be its limit. Now as $(1-tf)g_n = 1-(tf)^{n+1}$ for each n, $(1-tf)g_n$ converges to 1 in $n \geq \infty$. Therefore we have $(1-tf)\cdot g = 1$ in $K[[t]]$. This implies that $1-tf$ is invertible in $K[[t]]$, therefore as its inverse the rational expression $1/(1-tf)$ is well-defined in $K[[t]]$, and $1/(1-tf)$ is equal to the element $\sum_{n\geq 0}(tf)^n \in K[[t]]$ with the form of sum of geometric series.

We note that for the case of two or more variables, we can also define the formal power series ring $K[[t_1, \ldots, t_n]]$ recursively by applying the aforementioned construction to the rational function field in variable t_n over the quotient field of the integral domain $K[[t_1, \ldots, t_{n-1}]]$ as the coefficient field.

From now on, we introduce some classes of sequences satisfying recurrence relations with certain good properties and the corresponding classes of formal power series, and investigate their relations. We refer to Chapter 6 of [6] for more details. Here we focus on the case where $K = \mathbb{C}$.

Definition 2.2. We say that an element $f(x)$ of $\mathbb{C}[[x]]$ is *algebraic* if $f(x)$ is an algebraic element over the field $\mathbb{C}(x)$, that is, there exist $d \geq 1$ and $P_0, \ldots, P_d \in \mathbb{C}[x]$ satisfying

$$P_d(x)f(x)^d + \cdots + P_1(x)f(x) + P_0(x) = 0\,,\ P_d \neq 0\ .$$

The following property is deduced from a general theory on algebraic elements over some field; we omit the proof here.

Proposition 2.3. *The set of algebraic elements of* $\mathbb{C}[[x]]$ *is a subring of* $\mathbb{C}[[x]]$ *containing* $\mathbb{C}(x) \cap \mathbb{C}[[x]]$.

For example, the generating function $1/(1 - x - x^2)$ of Fibonacci numbers and the generating function $(1 - \sqrt{1 - 4x})/2x$ of Catalan numbers are algebraic (where $\sqrt{1 - 4x}$ means a square root of $1 - 4x$ which was written as $F_{1/2}(-4x)$ in the argument above).

We define the formal derivative for elements of $\mathbb{C}[[x]]$ by $\left(\sum_{n\geq 0} a_n x^n\right)' := \sum_{n\geq 1} n a_n x^{n-1}$. This satisfies the ordinary formulae for derivatives of addition and multiplication: $(f + g)' = f' + g'$ and $(fg)' = f'g + fg'$.

Definition 2.3. We say that an element $f(x)$ of $\mathbb{C}[[x]]$ is *differentially finite* (or shortly, *D-finite*) if the following holds as a linear spece over $\mathbb{C}(x)$:

$$\dim \mathsf{span}_{\mathbb{C}(x)} \left\{ \left(\frac{d}{dx}\right)^k f \mid k \geq 0 \right\} < \infty\ .$$

We note that the condition above is equivalent to that for some $d \geq 0$ and $P_0, \ldots, P_d \in \mathbb{C}[x]$ we have

$$P_d(x)f^{(d)}(x) + \cdots + P_1(x)f'(x) + P_0(x)f(x) = 0\,,\ P_d \neq 0\ . \tag{2}$$

Theorem 2.2. *For* $f(x) = \sum_{n \geq 0} a_n x^n \in \mathbb{C}[[x]]$*, the following conditions are equivalent:*

1. *f is D-finite.*
2. *For some $d \geq 0$ and $P_0, \ldots, P_d, Q \in \mathbb{C}[x]$ we have*

$$P_d(x)f^{(d)}(x) + \cdots + P_1(x)f'(x) + P_0(x)f(x) = Q(x)\,,\ P_d \neq 0\ .$$

3. *For some $d \geq 0$ and $P_0, \ldots, P_d \in \mathbb{C}[x]$ ($P_d \neq 0$) we have*

$$P_0(n)a_n + P_1(n)a_{n+1} + \cdots + P_d(n)a_{n+d} = 0$$

for any $n \geq 0$.

Proof. [1 $\Rightarrow$ 3] Starting from Eq.(2), we write the maximum of degrees of P_i's as D and for each $0 \leq k \leq D$, write the coefficient of x^k in P_i as $p_{i,k}$. Then the coefficient of x^n in the left-hand side of Eq.(2) is

$$\begin{aligned}
&\sum_{i=0}^{d} \sum_{j=0}^{D} p_{i,j} \cdot [\text{coeff. of } x^{n-j} \text{ in } f^{(i)}] \\
&= \sum_{i=0}^{d} \sum_{j=0}^{D} p_{i,j} \cdot (n-j+i)(n-j+i-1)\cdots(n-j+1)a_{n-j+i}\ ,
\end{aligned}$$

which is constantly equal to 0. Here we set $a_k := 0$ when $k < 0$. Therefore, by putting, for each $-D \leq k \leq d$,

$$R_k(n) := \sum_{\substack{0 \leq i \leq d,\, 0 \leq j \leq D \\ i-j=k}} p_{i,j} \cdot (n-j+i)(n-j+i-1)\cdots(n-j+1)\ ,$$

it is a polynomial in n and satisfies that for any $n \geq 0$ we have

$$\sum_{k=-D}^{d} R_k(n) a_{n+k} = 0 \ .$$

Now as $P_d \neq 0$, we have $p_{d,j_0} \neq 0$ for at least one j_0. For this j_0, $R_{d-j_0}(n)$ is of the folowing form

$$p_{d,j_0} \cdot (n-j_0+d) \cdots (n-j_0+1) + [\text{polynomial in } n \text{ of degree at most } d-1] \ ,$$

therefore $R_{d-j_0} \neq 0$. This implies that at least one of R_k's is non-zero, therefore there exists a d' with $-D \leq d' \leq d$ satisfying that for any $n \geq 0$ we have

$$\sum_{k=-D}^{d'} R_k(n) a_{n+k} = 0 \, , \ R_{d'} \neq 0 \ .$$

By putting $\widehat{R}_k(n) := R_{k-D}(n)$ for each $0 \leq k \leq d'+D$ and putting $n := m+D$ for each $m \geq 0$, the relation above becomes

$$\sum_{k=-D}^{d'} \widehat{R}_{k+D}(m+D) a_{m+k+D} = \sum_{k'=0}^{d'+D} \widehat{R}_{k'}(m+D) a_{m+k'} = 0 \ .$$

This gives a relation in Condition 3, as the $\widehat{R}_{k'}(m+D)$'s are polynomials in m and $\widehat{R}_{d'+D}(m+D)$ is not the zero polynomial.

[$3 \Rightarrow 2$] For each $0 \leq \ell \leq d$, as $(x+\ell)(x+\ell-1) \cdots (x+\ell-j+1)$ ($j \geq 0$) is a polynomial of degree j with the leading coefficient being 1, it follows that any polynomial of degree at most D can be expressed as a linear combination of such polynomials above over $j = 0, \ldots, D$. Based on this, for the relation in Condition 3, we write the maximum of degrees of the P_i's as D and write

$$P_\ell(x) = \sum_{j=0}^{D} b_{\ell,j} (x+\ell)(x+\ell-1) \cdots (x+\ell-j+1) \, , \ b_{\ell,j} \in \mathbb{C} \ .$$

Moreover, we set

$$Q(x) := \sum_{\ell=0}^{d} \sum_{j=0}^{D} b_{\ell,j} x^{j+d-\ell} f^{(j)}(x) \ .$$

Now when $m \geq d + D$, the coefficient of x^m in $Q(x)$ is

$$\begin{aligned}
&\sum_{\ell=0}^{d} \sum_{j=0}^{D} b_{\ell,j} \cdot [\text{coeff. of } x^{m-j-d+\ell} \text{ in } f^{(j)}] \\
&= \sum_{\ell=0}^{d} \sum_{j=0}^{D} b_{\ell,j} \cdot (m-d+\ell)(m-d+\ell-1) \cdots (m-d+\ell-j+1) a_{m-d+\ell} \\
&= \sum_{\ell=0}^{d} P_\ell(m-d) a_{m-d+\ell} \ .
\end{aligned}$$

The right-hand side is equal to the left-hand side of the relation in Condition 3 with $n = m - d$, which is 0. Hence $Q(x)$ is a polynomial of degree at most $d + D - 1$. Moreover, as $P_d \neq 0$, we have $b_{d,j_0} \neq 0$ for some j_0. Now the coefficient of $f^{(j_0)}(x)$ in the right-hand side of the definition of $Q(x)$ is

$$b_{d,j_0} x^{j_0} + [\text{monomials in } x \text{ of degrees at least } j_0 + 1] \ ,$$

which is not the zero polynomial. Hence by rearranging the defining equality for $Q(x) \in \mathbb{C}[x]$, we obtain a relation as in Condition 2.

$[2 \Rightarrow 1]$ By differentiating both sides of the relation in Condition 2 $\deg Q +$ 1 times, the right-hand side becomes 0, while the left-hand side becomes $R_{d+\deg Q}(x) f^{(d+\deg Q)}(x) + \cdots + R_1(x) f'(x) + R_0(x) f(x)$ for some polynomials $R_0, \ldots, R_{d+\deg Q}$. Combining this and the property $R_{d+\deg Q}(x) = P_d(x) \neq 0$ yields a relation as in Eq.(2). Hence the claim holds. □

By this theorem, the class of D-finite formal power series can be regarded as the class of generating functions of sequences having recurrence relations of the form of linear combination with coefficients being polynomials in n.

Similarly to the case of algebraic formal power series, the following property holds.

Proposition 2.4. *The set of D-finite elements of* $\mathbb{C}[[x]]$ *is a subring of* $\mathbb{C}[[x]]$ *containing* $\mathbb{C}(x) \cap \mathbb{C}[[x]]$.

Proof. For the claim that the set in the statement contains $\mathbb{C}(x)$, it can be deduced by setting $d = 0$ in Condition 2 of Theorem 2.2. For the remaining claim, let $f, g \in \mathbb{C}[[x]]$ be D-finite elements. By definition, we can take finite-dimensional $\mathbb{C}(x)$-linear spaces V_f, V_g satisfying that all $f^{(n)}$'s belong to V_f, all $g^{(n)}$'s belong to V_g, and each of V_f and V_g is generated by some finite subset of $\mathbb{C}[[x]]$. Now for $n \geq 0$, we have that $(f \pm g)^{(n)} \in V_f + V_g$ and $V_f + V_g$ is finite-dimensional, therefore $f \pm g$ is also D-finite. On the other hand, when $\{\alpha_1, \ldots, \alpha_{d_f}\} \subseteq \mathbb{C}[[x]]$ is a finite generating set of V_f and $\{\beta_1, \ldots, \beta_{d_g}\} \subseteq \mathbb{C}[[x]]$ is a finite generating set of V_g, the $\mathbb{C}(x)$-linear space W generated by the elements $\alpha_i \beta_j$ $(1 \leq i \leq d_f,\ 1 \leq j \leq d_g)$ is also finite-dimensional and satisfies that for any $n, m \geq 0$ we have $f^{(n)} g^{(m)} \in W$. Now for $n \geq 0$, we have

$$(fg)^{(n)} = \sum_{k=0}^{n} \binom{n}{k} f^{(k)} g^{(n-k)} \in W \ .$$

Therefore fg is also D-finite. Hence the claim holds. □

For those two classes (algebraic and D-finite ones), we have the following relation.

Theorem 2.3. *If $f \in \mathbb{C}[[x]]$ is algebraic, then f is D-finite.*

Proof. We note that as f is algebraic, the field $\mathbb{C}(x, f)$ obtained by appending f to $\mathbb{C}(x)$ is finite-dimensional over $\mathbb{C}(x)$. Take a relation as follows yielded by the fact that f is algebraic,

$$P_d f^d + \cdots + P_1 f + P_0 = 0 \,,\ P_i \in \mathbb{C}[x] \,,\ P_d \neq 0$$

in a way that d is minimal. Differentiating both sides yields

$$\sum_{i=0}^{d} P_i' f^i + \left(\sum_{i=1}^{d} P_i \cdot i f^{i-1} \right) \cdot f' = 0 \ .$$

By the minimality of d, the coefficient of f' is non-zero, therefore by solving this equation in f' we have $f' \in \mathbb{C}(x, f)$. Now if $f^{(n)} \in \mathbb{C}(x, f)$, then the

formula for derivative of fraction implies that $f^{(n+1)}$ can be written as a rational expression of x, f, and f', therefore by the fact $f' \in \mathbb{C}(x, f)$ we have $f^{(n+1)} \in \mathbb{C}(x, f)$. By using it recursively, it follows that for any $n \geq 0$, $f^{(n)}$ belongs to the finite-dimensional $\mathbb{C}(x)$-linear space $\mathbb{C}(x, f)$. Hence f is D-finite and the claim holds. □

By Theorem 2.3, for example, the generating function of Catalan numbers is D-finite (which would be not obvious from the original definition of Catalan numbers). We note that the converse of Theorem 2.3 does not hold (see Chapter 6, Exercise 6.1 of [6]).

We show that if two sequences have D-finite generating functions, then the component-wise product of those sequences also has a D-finite generating function. For the purpose, we prepare some definitions and properties.

Definition 2.4. We define an equivalence relation $\sim$ over the set of $\mathbb{C}$-valued sequences by

$(a_n)_n \sim (b_n)_n \overset{\text{def}}{\Leftrightarrow}$ there exists an N satisfying that for any $m \geq N$ we have $a_m = b_m$.

We call its equivalence class $[a_n]_n$ the *germ* involving $(a_n)_n$.

Note that the component-wise addition and multiplication for sequences naturally induce addition and multiplication for germs. It also holds that the property "it has a D-finite generating function" for sequences is preserved by this equivalence relation. Namely, we have the following result.

Proposition 2.5. *Suppose that sequences $(a_n)_n$ and $(b_n)_n$ are equivalent in the sense above, and let $f(x)$ and $g(x)$ be generating functions of $(a_n)_n$ and $(b_n)_n$, respectively. If f is D-finite, then g is also D-finite.*

Proof. By the hypothesis, we can take a finite-dimensional $\mathbb{C}(x)$-linear space V involving all $f^{(n)}$'s. Moreover, as $(a_n)_n \sim (b_n)_n$, we have $g = f + P$ for some polynomial $P(x)$. Now for any $n \geq 0$ we have $g^{(n)} = f^{(n)} + P^{(n)} \in V + \mathbb{C}(x)$. As $V + \mathbb{C}(x)$ is also finite-dimensional over $\mathbb{C}(x)$, it follows that g is D-finite. Hence the claim holds. □

Lemma 2.3. *For the generating function $f(x)$ of a sequence $(a_n)_n$, the following conditions are equivalent.*

1. *f is D-finite.*
2. *The set $\{[a_{n+i}]_n \mid i \geq 0\}$ generates a finite-dimensional linear space over $\mathbb{C}(n)$.*

Proof. [1 $\Rightarrow$ 2] Dividing the relation in Condition 3 of Theorem 2.2 by P_d and rearranging it give a relation of the form

$$a_{n+d} = Q_0(n)a_n + Q_1(n)a_{n+1} + \cdots + Q_{d-1}(n)a_{n+d-1} \ , \ Q_i \in \mathbb{C}(x) \ .$$

For any $k \geq 0$, by substituting $n+k$ to the n in the relation above, for any $n \geq 0$ we have

$$a_{n+d+k} = Q_0(n+k)a_{n+k} + Q_1(n+k)a_{n+k+1} + \cdots + Q_{d-1}(n+k)a_{n+k+d-1} \ .$$

By translating it to the case of germs, for any $m \geq d$, $[a_{n+m}]_n$ can be written as a linear combination of $[a_n]_n, [a_{n+1}]_n, \ldots, [a_{n+m-1}]_n$ over $\mathbb{C}(n)$. By using it recursively, it follows that any $[a_{n+m}]_n$ belongs to the finite-dimensional $\mathbb{C}(x)$-linear space generated by $[a_n]_n, [a_{n+1}]_n, \ldots, [a_{n+d-1}]_n$. Hence Condition 2 holds.

[2 $\Rightarrow$ 1] By the hypothesis, there exist $d \geq 0$ and $P_0, \ldots, P_d \in \mathbb{C}[x]$ ($P_d \neq 0$) for which the following relation for germs holds:

$$P_0(n)[a_n]_n + \cdots + P_d(n)[a_{n+d}]_n = [0] \ .$$

Therefore, we can take an N for which for any $n \geq N$ we have

$$P_0(n)a_n + \cdots + P_d(n)a_{n+d} = 0 \ .$$

By putting $R(x) := x(x-1)\cdots(x-N+1)$, for any $n \geq N$ we have

$$P_0(n)R(n)a_n + \cdots + P_d(n)R(n)a_{n+d} = 0 \ .$$

This relation also holds for $0 \leq n \leq N-1$, as now $R(n) = 0$. As $P_d R \neq 0$, this means that Condition 3 in Theorem 2.2 is satisfied for $(a_n)_n$. Hence f is D-finite and the claim holds. $\square$

Theorem 2.4. *Let $f(x)$ and $g(x)$ be generating functions of sequences $(a_n)_n$ and $(b_n)_n$, respectively, and let $f*g$ denote the generating function of sequence $(a_n b_n)_n$. If f and g are D-finite, then $f * g$ is also D-finite.*

Proof. Let V_a and V_b be the $\mathbb{C}(x)$-linear spaces generated by germs $[a_{n+i}]_n$ ($i \geq 0$) and germs $[b_{n+i}]_n$ ($i \geq 0$), respectively. They are both finite-dimensional by Lemma 2.3. Let $(\alpha_1, \ldots, \alpha_{d_a})$ and $(\beta_1, \ldots, \beta_{d_b})$ be bases of V_a and V_b, respectively. Then the $\mathbb{C}(x)$-linear space W generated by the elements $\alpha_i \beta_j$ ($1 \leq i \leq d_a$, $1 \leq j \leq d_b$) is also finite-dimensional, and for any $i \geq 0$ we have $[a_{n+i} b_{n+i}]_n = [a_{n+i}]_n \cdot [b_{n+i}]_n \in W$. Hence by Lemma 2.3, $f * g$ is also D-finite. □

At the end of this section, we mention the following fact about the generating function of "diagonal part" of a multi-dimensional sequence. We omit the proof of this fact. See Chapter 6, Exercise 6.61 of [6] for the former claim and Chapter 6, Theorem 6.3.3 of [6] for the latter claim.

Theorem 2.5. *For an n-variate formal power series*

$$f(x_1, \ldots, x_n) = \sum_{k_1, \ldots, k_n \geq 0} a_{k_1, \ldots, k_n} x_1^{k_1} \cdots x_n^{k_n} \in \mathbb{C}[[x_1, \ldots, x_n]]$$

we define $\mathrm{diag}(f)(x) := \sum_{k \geq 0} a_{k, \ldots, k} x^k$. *If this f is also an element of $\mathbb{C}(x_1, \ldots, x_n)$, then* $\mathrm{diag}(f)(x)$ *is D-finite. If moreover $n = 2$, then* $\mathrm{diag}(f)(x)$ *is algebraic.*

Exercises

Problem 1. Here we write the Fibonacci numbers as $(F_n)_n$, and we define the Lucas number $(L_n)_n$ by $L_0 = 2$, $L_1 = 1$, and a recurrence relation $L_{n+2} = L_n + L_{n+1}$. Prove, by using generating functions, that the relation $L_n = 2F_n - F_{n-1}$ ($n \geq 1$) holds.

(Hint: It is now not necessary to determine the coefficients of generating functions of Fibonacci and Lucas numbers; it is sufficient for our present purpose to obtain their rational expressions.)

Problem 2. For an integer $n \geq 1$, we write the Young diagram of partition (n, n) as Y. We consider to fill different numbers from 1 to $2n$ into the boxes of Y, one per each box. We call such an arrangement a *standard Young tableau* of shape Y if the following conditions hold.

- The numbers in each row is increasing from left to right.
- The numbers in each column is increasing from top to bottom.

Prove that the number of standard Young tableaux of shape Y is equal to C_n, by constructing a bijection between the set of Dyck paths from $(0, 0)$ to $(2n, 0)$ and the set of standard Young tableaux of shape Y.

Problem 3. For integer $n \geq 0$, let a_n denote the number of ways of expressing (n, n) as a sum of vectors $(1, 0)$, $(0, 1)$, and $(1, 1)$ that can be used multiple times (where we distinguish the sums that only differ in the order of vectors). Prove that its generating function $\sum_{n \geq 0} a_n x^n$ is algebraic.

(Hint: Apply the latter part of Theorem 2.5, by appropriately constructing a two-dimensional sequence $(b_{n,m})_{n,m}$ in a way that its generating function is a rational function and $a_n = b_{n,n}$.)

Problem 4. For Catalan numbers C_n, let $D_n := C_n^2$. Give a relation for the sequence $(D_n)_n$ as in Condition 3 of Theorem 2.2.

(Comment: The generating function of Catalan numbers is D-finite as mentioned above. Therefore the generating function of $(D_n)_n$ is also D-finite by Theorem 2.4; hence the existence of such a relation as in the statement is guaranteed.)

3 Principle of Inclusion-Exclusion and Möbius Inversion Formula

This section is based on Chapter 3 of Stanley's book [5].

We say that an element $\sigma \in S_n$ of the symmetric group of degree n is a *derangement* if σ has no fixed points, i.e., for any $a \in \{1, \ldots, n\}$ we have $\sigma(a) \neq a$. Here we write the set of derangements on n letters as D_n. For example, we set $n = 3$ and try to count the elements of D_3. The total number of elements of S_3 is

$$3! \ ,$$

but this involves some elements to be excluded, i.e., ones fixing 1, ones fixing 2, and ones fixing 3. The number of each of them is 2! (as it is equal to the number of permutations on 2 elements); simply substituting them yields

$$3! - 3 \cdot 2! \ .$$

But now, for example, focusing on elements fixing both 1 and 2, the formula above counts them $1 - 2 = -1$ time, which is too few. To cancel it, we add the number 1! of such elements (equal to the number of permutations on 1 element). Doing similarly for elements fixing both 1 and 3 and for elements fixing both 2 and 3, we have the count

$$3! - 3 \cdot 2! + 3 \cdot 1! \ .$$

However, now the elements fixing all of 1, 2, and 3 (permutations on 0 element) are counted $1 - 3 + 3 = 1$ time; to cancel it, we have to subtract the number 0! of such elements. As a result, we have

$$|D_3| = 3! - 3 \cdot 2! + 3 \cdot 1! - 0! = 6 - 3 \cdot 2 + 3 \cdot 1 - 1 = 2$$

(indeed, we have $D_3 = \{(123), (132)\}$).

As a generalization of this argument, by adjusting the additions and subtractions carefully, we have the formula

$$\begin{aligned} |D_n| &= n! - \binom{n}{1} \cdot (n-1)! + \binom{n}{2} \cdot (n-2)! - \cdots + (-1)^n \cdot 0! \\ &= \sum_{k=0}^{n} (-1)^k \frac{n!}{k!} = n! \cdot \sum_{k=0}^{n} \frac{(-1)^k}{k!} \approx \frac{n!}{e} \ . \end{aligned}$$

From now on, we explain a theory called *Principle of Inclusion-Exclusion* that can formalize the part "adjusting the additions and subtractions carefully" above.

In general, the Principle of Inclusion-Exclusion is described by using "Möbius functions" on partially ordered sets.

Definition 3.1. Let P be a set and let $\preceq$ be a binary relation on P. We say that $(P, \preceq)$ (or simply, P) is a *partially ordered set* (*poset*) if the following conditions hold. We also call such $\preceq$ a *partial order*.

1. For any $x \in P$, we have $x \preceq x$ (reflexivity).
2. For any $x, y \in P$, if $x \preceq y$ and $y \preceq x$ then we have $x = y$ (antisymmetry).
3. For any $x, y, z \in P$, if $x \preceq y$ and $y \preceq z$ then we have $x \preceq z$ (transitivity).

If moreover the following condition holds:

- For any $x, y \in P$, we have $x \preceq y$ or $y \preceq x$,

then we say that P is a *totally ordered set* and $\preceq$ is a *total order*.

From now on, when we use a symbol like $\preceq$ to express an order, we use a symbol like $x \prec y$ to mean "$x \preceq y$ and $x \neq y$".

Example 3.1. For elements x, y of a poset P, we say that y *covers* x and x is *covered by* y if we have $x \prec y$ and there exists no z with $x \prec z \prec y$. A *Hasse diagram* of P is given by regarding elements of P as points and, if y covers x, then joining x and y by a line where y is placed above x. For example, the set $\mathbb{Z}_{\geq 0}$ of non-negative integers forms a totally ordered set with the usual large/small relation as the order, for which (a part of) a Hasse diagram is as in the left part of Figure 4. (We note that by the transitivity, an element in the diagram and some other element that can be reached from the original element by following some upward lines have an order relation even if they are not directly joined by a line.) Similarly, the set of all subsets of $\{1, 2, 3\}$, ordered by the inclusion relation $\subseteq$, forms a poset with Hasse diagram as in the middle part of Figure 4, and the set of Young diagrams, ordered by the inclusion relation (where all Young diagrams are upper-left aligned), forms a poset with (a part of) Hasse diagram as in the right part of Figure 4.

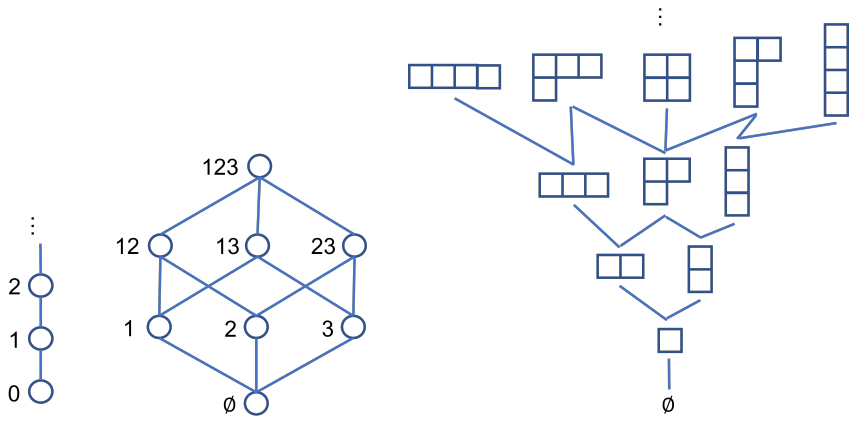


Figure 4: Examples of Hasse diagrams for posets

Definition 3.2. Let $(P, \preceq_P)$ and $(Q, \preceq_Q)$ be posets, and let $f\colon P \to Q$. We say that f is *order-preserving* if for any $x, y \in P$ with $x \preceq_P y$ we have $f(x) \preceq_Q f(y)$. We say that f is an *isomorphism* if f is bijective and both

f and f^{-1} are order-preserving. When such an isomorphism exists, we say that P and Q are *isomorphic* and write $P \simeq Q$.

Definition 3.3. Let P be a poset. For $x, y \in P$ with $x \preceq y$, we define

$$[x,y]_P := \{z \in P \mid x \preceq z \preceq y\} \ .$$

Under the notation, we say that P is *locally finite* if for any $x, y \in P$ with $x \preceq y$, $[x,y]_P$ is always a finite set.

For any locally finite poset P, we define

$$I(P) := \{(x,y) \in P^2 \mid x \preceq y\} \ ,$$

and define the addition and multiplication on the set $\mathbb{C}^{I(P)} := \{f \colon I(P) \to \mathbb{C}\}$ as follows: for any $f, g \in \mathbb{C}^{I(P)}$ and $(x,y) \in I(P)$,

$$(f+g)(x,y) := f(x,y) + g(x,y) \ , \ (fg)(x,y) := \sum_{\substack{z \ ; \\ x \preceq z \preceq y}} f(x,z)g(z,y) \ .$$

We define the scalar multiplication on $\mathbb{C}^{I(P)}$ as follows: for any $f \in \mathbb{C}^{I(P)}$, $\alpha \in \mathbb{C}$, and $(x,y) \in I(P)$,

$$(\alpha \cdot f)(x,y) := \alpha f(x,y) \ .$$

On the other hand, we define the addition, multiplication, and scalar multiplication on the set $\mathbb{C}^P := \{f \colon P \to \mathbb{C}\}$ by $(f+g)(x) := f(x) + g(x)$, $(fg)(x) := f(x)g(x)$, and $(\alpha \cdot f)(x) := \alpha f(x)$. Then it can be straightforwardly checked that $\mathbb{C}^P$ forms a $\mathbb{C}$-algebra.

Proposition 3.1. *Under those definitions, $\mathbb{C}^{I(P)}$ is a $\mathbb{C}$-algebra. Let $\delta \colon I(P) \to \mathbb{C}$ be defined by*

$$\delta(x,y) = \begin{cases} 1 & \textit{(if } x = y\textit{)} \\ 0 & \textit{(otherwise).} \end{cases}$$

Then δ is the multiplicative identity element of $\mathbb{C}^{I(P)}$. Moreover, if for any $f \in \mathbb{C}^P$ we define $\widehat{f} \in \mathbb{C}^{I(P)}$ by $\widehat{f}(x,y) := f(x)\delta(x,y)$, the correspondence $f \mapsto \widehat{f}$ forms a homomorphism as $\mathbb{C}$-algebras from $\mathbb{C}^P$ to $\mathbb{C}^{I(P)}$.

Proof. It is obvious by definition that $\mathbb{C}^{I(P)}$ forms a $\mathbb{C}$-linear space.

The associativity for multiplication holds as follows:

$$\begin{aligned}((fg)h)(x,y) &= \sum_{x \preceq z \preceq y} (fg)(x,z) \cdot h(z,y) \\ &= \sum_{x \preceq w \preceq z \preceq y} f(x,w) g(w,z) \cdot h(z,y) \\ &= \sum_{x \preceq w \preceq y} f(x,w) \cdot (gh)(w,y) = (f(gh))(x,y) \ .\end{aligned}$$

The distributive law (from the right) holds as follows:

$$\begin{aligned}((f+g)h)(x,y) &= \sum_{x \preceq z \preceq y} (f+g)(x,z) \cdot h(z,y) \\ &= \sum_{x \preceq z \preceq y} (f(x,z) + g(x,z)) \cdot h(z,y) \\ &= \sum_{x \preceq z \preceq y} (f(x,z)h(z,y) + g(x,z)h(z,y)) \\ &= (fh)(x,y) + (gh)(x,y) = (fh+gh)(x,y) \ .\end{aligned}$$

The case of the other side is similar.

For any $f, g \in \mathbb{C}^{I(P)}$, $\alpha \in \mathbb{C}$, and $(x,y) \in I(P)$, we have

$$\begin{aligned}(\alpha \cdot (fg))(x,y) &= \alpha (fg)(x,y) \\ &= \alpha \sum_{x \preceq z \preceq y} f(x,z) g(x,y) \\ &= \sum_{x \preceq z \preceq y} (\alpha f(x,z)) g(x,y) \\ &= \sum_{x \preceq z \preceq y} (\alpha \cdot f)(x,z) g(x,y) = ((\alpha \cdot f)g)(x,y) \ ,\end{aligned}$$

therefore we have $\alpha \cdot (fg) = (\alpha \cdot f)g$ and it holds similarly that $\alpha \cdot (fg) = f(\alpha \cdot g)$. Summarizing, $\mathbb{C}^{I(P)}$ is a $\mathbb{C}$-algebra. Moreover, as

$$(\delta f)(x,y) = \sum_{x \preceq z \preceq y} \delta(x,z) f(z,y) = 1 \cdot f(x,y) = f(x,y) \ ,$$

we have $\delta f = f$ and similarly $f\delta = f$. Therefore δ is the multiplicative identity element of $\mathbb{C}^{I(P)}$.

For any $f, g \in \mathbb{C}^P$, $\alpha \in \mathbb{C}$, and $(x, y) \in I(P)$, we have

$$\begin{aligned}
(\widehat{f} + \widehat{g})(x,y) &= \widehat{f}(x,y) + \widehat{g}(x,y) \\
&= f(x)\delta(x,y) + g(x)\delta(x,y) \\
&= (f(x) + g(x))\delta(x,y) \\
&= (f+g)(x)\delta(x,y) = \widehat{f+g}(x,y) \ , \\
(\widehat{f}\widehat{g})(x,y) &= \sum_{x \preceq z \preceq y} \widehat{f}(x,z)\widehat{g}(z,y) \\
&= \sum_{x \preceq z \preceq y} f(x)\delta(x,z)g(z)\delta(z,y) \\
&= f(x)g(x)\delta(x,y) \\
&= (fg)(x)\delta(x,y) = \widehat{fg}(x,y) \ , \\
(\alpha \cdot \widehat{f})(x,y) &= \alpha \widehat{f}(x,y) \\
&= \alpha f(x)\delta(x,y) \\
&= (\alpha \cdot f)(x)\delta(x,y) = \widehat{\alpha \cdot f}(x,y) \ .
\end{aligned}$$

Therefore the correspondence $f \mapsto \widehat{f}$ is a homomorphism of $\mathbb{C}$-algebras. Hence the claim holds. □

From now on, we suppose that for any $x \in P$ the set $\wedge^x := \{w \in P \mid w \preceq x\}$ is always finite. For any $f \in \mathbb{C}^{I(P)}$ and $\varphi \in \mathbb{C}^P$, we define $\varphi * f \in \mathbb{C}^P$ by

$$(\varphi * f)(x) := \sum_{w \preceq x} \varphi(w) f(w,x) \ .$$

Lemma 3.1. *For any $f, g \in \mathbb{C}^{I(P)}$ and $\varphi \in \mathbb{C}^P$, we have $\varphi * (fg) = (\varphi * f) * g$ and $\varphi * \delta = \varphi$.*

Proof. For any $x \in P$, we have

$$\begin{aligned}(\varphi * (fg))(x) &= \sum_{w \preceq x} \varphi(w)(fg)(w,x) \\ &= \sum_{w \preceq z \preceq x} \varphi(w) f(w,z) g(z,x) \\ &= \sum_{z \preceq x} (\varphi * f)(z) g(z,x) = ((\varphi * f) * g)(x) \ ,\end{aligned}$$

$$(\varphi * \delta)(x) = \sum_{w \preceq x} \varphi(w)\delta(w,x) = \varphi(x)\delta(x,x) = \varphi(x) \ .$$

Hence the claim holds. □

Definition 3.4. Let P be a locally finite poset. We define the *Möbius function* $\mu = \mu_P \in \mathbb{C}^{I(P)}$ on P recursively as follows: for $x \in P$ define $\mu(x,x) := 1$, and for $x \prec y$ define $\mu(x,y) := -\sum_{z\,;\, x \prec z \preceq y} \mu(z,y)$.

From now on, let P be a locally finite poset.

Lemma 3.2. *Let $\mathbb{1} \in \mathbb{C}^{I(P)}$ be defined by $\mathbb{1}(x,y) = 1$ for any $(x,y) \in I(P)$. Then we have $\mathbb{1}\mu = \delta$.*

Proof. For any $x \in P$, we have

$$(\mathbb{1}\mu)(x,x) = \mathbb{1}(x,x)\mu(x,x) = 1 = \delta(x,x) \ .$$

For any $x \prec y$, we have

$$\begin{aligned}(\mathbb{1}\mu)(x,y) &= \sum_{x \preceq z \preceq y} \mathbb{1}(x,z)\mu(z,y) \\ &= \sum_{x \preceq z \preceq y} \mu(z,y) \\ &= \mu(x,y) + \sum_{x \prec z \preceq y} \mu(z,y) = 0 = \delta(x,y) \ .\end{aligned}$$

Hence the claim holds. □

Lemma 3.3. *We have* $\mu\mathbb{1} = \delta$.

Proof. The claim is equivalent to that for any $(x, y) \in I(P)$ we have $\sum_{x \preceq z \preceq y} \mu(x, z) = \delta(x, y)$. We prove the latter property by mathematical induction with respect to the number of edges in the longest upward path from x to y in a Hasse diagram of P. When $x = y$, the claim holds as $\mu(x, x) = 1 = \delta(x, x)$. From now, we suppose that $x \prec y$. By the recursive definition of μ, we have

$$\sum_{x \preceq z \preceq y} \mu(x, z) = 1 + \sum_{x \prec z \preceq y} \mu(x, z) = 1 - \sum_{x \prec z \preceq y} \sum_{x \prec w \preceq z} \mu(w, z) = 1 - \sum_{x \prec w \preceq y} \sum_{w \preceq z \preceq y} \mu(w, z) \ .$$

By the induction hypothesis, for any $w \in [x, y]_P \backslash \{x\}$ we have $\sum_{w \preceq z \preceq y} \mu(w, z) = \delta(w, y)$. Therefore we have

$$\sum_{x \preceq z \preceq y} \mu(x, z) = 1 - \sum_{x \prec w \preceq y} \delta(w, y) = 1 - \delta(y, y) = 1 - 1 = 0 = \delta(x, y) \ .$$

Hence the claim holds. □

Theorem 3.1. *For any* $f, g \in \mathbb{C}^P$, *the conditions* $g = f * \mathbb{1}$ *and* $f = g * \mu$ *are equivalent. These properties are called* Möbius inversion formula.

Proof. By Lemmas 3.2 and 3.3, μ is the inverse of $\mathbb{1}$ (from both sides). Therefore, if $g = f * \mathbb{1}$ then we have

$$g * \mu = (f * \mathbb{1}) * \mu = f * (\mathbb{1}\mu) = f * \delta = f \ ,$$

while if $f = g * \mu$ then we have

$$f * \mathbb{1} = (g * \mu) * \mathbb{1} = g * (\mu\mathbb{1}) = g * \delta = g \ .$$

Hence the claim holds. □

As an application of Möbius inversion formula, we revisit the aforementioned example of derangements. For each $T \subseteq \{1, \ldots, n\}$, we define

$$\mathcal{F}(T) := \{\sigma \in S_n \mid \sigma(a) \neq a \Leftrightarrow a \in T\} \, , \ f(T) := |\mathcal{F}(T)| \ ,$$

$$\mathcal{G}(T) := \{\sigma \in S_n \mid \sigma(a) \neq a \Rightarrow a \in T\}\,,\ g(T) := |\mathcal{G}(T)|\ .$$

We have $D_n = \mathcal{F}(\{1,\dots,n\})$ by definition. Now $\mathcal{G}(T)$ is the disjoint union of $\mathcal{F}(U)$'s over $U \subseteq T$, therefore we have $g(T) = \sum_{U\subseteq T} f(U)$. We consider a poset P that is the set of subsets of $\{1,\dots,n\}$ ordered by the inclusion relation. Then the relation above becomes

$$g(T) = \sum_{U\subseteq T} f(U) = \sum_{U\subseteq T} f(U)\mathbb{1}(U,T) = (f * \mathbb{1})(T)\ ,$$

therefore $g = f * \mathbb{1}$. Now Möbius inversion formula implies that $f = g * \mu_P$, in particular

$$|D_n| = f(\{1,\dots,n\}) = \sum_{U\subseteq\{1,\dots,n\}} g(U)\mu_P(U,\{1,\dots,n\})\ .$$

Moreover, the Möbius function in this case is determined as follows.

Proposition 3.2. *For the P above, we have $\mu_P(I,J) = (-1)^{|J|-|I|}$.*

Proof. We use mathematical induction with respect to $|J| - |I|$. The case $I = J$ is obvious by definition. For the case $I \subsetneq J$, we have

$$\mu_P(I,J) = -\sum_{I\subsetneq K\subseteq J} \mu_P(K,J) = -\sum_{I\subsetneq K\subseteq J} (-1)^{|J|-|K|}$$

(here we used the induction hypothesis)

$$\begin{aligned}
&= -\sum_{k=1}^{|J|-|I|} \binom{|J|-|I|}{k} (-1)^{|J|-(|I|+k)} \\
&= (-1)^{|J|-|I|} - \sum_{k=0}^{|J|-|I|} \binom{|J|-|I|}{k} (-1)^{|J|-|I|-k} \\
&= (-1)^{|J|-|I|} - (1-1)^{|J|-|I|} = (-1)^{|J|-|I|}\ ,
\end{aligned}$$

therefore the claim holds in this case. Hence the claim holds. □

Substituting this result into the relation above implies

$$|D_n| = \sum_{U \subseteq \{1,\dots,n\}} (-1)^{n-|U|} g(U) \ .$$

Moreover, as $\mathcal{G}(U)$ is the set of permutations fixing any point not belonging to U, we have $g(U) = |U|!$. Summarizing, we have

$$\begin{aligned} |D_n| = \sum_{U \subseteq \{1,\dots,n\}} (-1)^{n-|U|} |U|! &= \sum_{k \geq 0} (-1)^{n-k} \binom{n}{k} k! \\ &= \sum_{k \geq 0} (-1)^{n-k} \frac{n!}{(n-k)!} = \sum_{k \geq 0} (-1)^k \frac{n!}{k!} \ , \end{aligned}$$

which is the fact mentioned above.

As another example, let P be the set of positive integers, and define the order relation in a way that for any $a, b \in P$ we have $a \preceq b$ if and only if $a \mid b$ (i.e., a divides b). Then P forms a poset. Now the Möbius function is determined as follows.

Proposition 3.3. *For the P above, consider a prime factorization $b/a = p_1^{e_1} \cdots p_\ell^{e_\ell}$ (where the p_i's are distinct primes and $e_i > 0$). If all the e_i's are 1, then we have $\mu_P(a,b) = (-1)^\ell$; otherwise we have $\mu_P(a,b) = 0$.*

Proof. We use mathematical induction with respect to $e_1 + \cdots + e_\ell$. When the e_i's are all 1, we define a map f from the set $\mathcal{P}(\{1,\dots,\ell\})$ of subsets of $\{1,\dots,\ell\}$ to $[a,b]_P$ by $f(\{i_1,\dots,i_k\}) := a p_{i_1} \cdots p_{i_k}$. Then this f is an isomorphism. By definition, the value of $\mu_P(a,b)$ depends solely on the order structure of $[a,b]_P$. Therefore the Möbius function on $\mathcal{P}(\{1,\dots,\ell\})$ studied above implies that $\mu_P(a,b) = (-1)^\ell$.

On the other hand, when some e_i is at least 2, by definition we have

$$\mu_P(a,b) = - \sum_{\substack{c \neq a\,; \\ a|c\,,\, c|b}} \mu_P(c,b) \ .$$

Now the induction hypothesis implies that when b/c has a square divisor we have $\mu_P(c,a) = 0$. Therefore the parameter c in the sum above essentially

varies over the c's for which b/c is square-free (note that b/a has a square divisor by the current assumption). This implies that

$$\mu_P(a,b) = -\sum_{\substack{c\,;\\ bp_1^{-1}\cdots p_\ell^{-1}|c\,,\,c|b}} \mu_P(c,b) = -\sum_{c\in[bp_1^{-1}\cdots p_\ell^{-1},b]_P} \mu_P(c,b) = 0 \ .$$

Hence the claim holds. □

For functions f and g in positive integer n, the condition that $g(n) = \sum_{d|n} f(d)$ always holds is equivalent to $g = f * \mathbb{1}$. Now Möbius inversion formula implies that $f = g * \mu_P$ and hence $f(n) = \sum_{d|n} g(d)\mu_P(d,n)$. This is nothing but the Möbius inversion formula in elementary number theory.

Exercises

Problem 1. Let $P = \{1, 2, \ldots, 10\}$, and define a partial order on P in a way that $a \preceq b$ if and only if a divides b. Draw a Hasse diagram of P.

Problem 2. Give an example of posets P, Q and a map $f\colon P \to Q$ for which f is bijective and order-preserving but is not an isomorphism.

(Comment: For example, for the case of groups, if $f\colon G \to H$ is a bijective homomorphism, then f is always an isomorphism. This problem says that a similar property does not hold for the case of posets.)

Problem 3. Let P, Q be posets and $f\colon P \to Q$ be a bijective order-preserving map. Prove that if moreover the order on P is a total order, then f is an isomorphism.

Problem 4. Let P be the set of Young diagrams with at most 3 boxes, ordered by the inclusion relation. Determine the value of $\mu_P(\emptyset, Y)$ for every $Y \in P$.

4 Ordered Sets and Lattices

This section is based on Chapter 3 of [5].

In the area of combinatorics (or discrete mathematics), posets themselves have also been a research topic, where several notions, subclasses consisting of special kinds of posets, etc. are studied.

Definition 4.1. Let P be a poset and let S be its subset. By restricting the order relation of P to S, S also forms a poset. This poset is called a *partially ordered subset* (*subposet*) of P. Moreover:

- We say that $x \in P$ is an *upper bound* of S if for any $y \in S$ we have $x \succeq y$. Similarly, we say that $x \in P$ is a *lower bound* of S if for any $y \in S$ we have $x \preceq y$.
- We say that $x \in S$ is *maximal* in S if for any $y \in S$ with $y \succeq x$ we have $y = x$. Similarly, we say that $x \in S$ is *minimal* in S if for any $y \in S$ with $y \preceq x$ we have $y = x$.
- We say that $x \in S$ is *maximum* in S if x itself is an upper bound of S. Similarly, we say that $x \in S$ is *minimum* in S if x itself is a lower bound of S.

When P has a maximum element (respectively, minimum element), we sometimes write this element as $\max P$, 1, or 1_P (respectively, $\min P$, 0, or 0_P).

Example 4.1. We consider the set $P = \{1, 2, \ldots, 10\}$ ordered by divisibility relation, i.e., $a \preceq b \Leftrightarrow a \mid b$. The only upper bound of $S_1 = \{2, 3\}$ is 6, and the only lower bound of S_1 is 1. Upper bounds of $S_2 = \{4, 6\}$ do not exist, and the only lower bounds of S_2 are 1 and 2. The maximal elements of P are $6, 7, 8, 9, 10$, and the maximum element of P does not exist. The minimum element of P is 1, and the only minimal element of P is also 1 (see the exercises).

Definition 4.2. Let P be a poset and let $S \subseteq P$. If the set of the upper bounds of S in P has the minimum element, then we call it the *join* of S, denoted by $\bigvee S$, $\bigvee_{x \in S} x$, etc. In the case where the S is a finite set $\{x_1, \ldots, x_n\}$, we also write the join of S as $x_1 \vee \cdots \vee x_n$. On the other hand, if the set of the lower bounds of S in P has the maximum element, then we call it the *meet* of S, denoted by $\bigwedge S$, $\bigwedge_{x \in S} x$, etc. In the case where the S is a finite set $\{x_1, \ldots, x_n\}$, we also write the meet of S as $x_1 \wedge \cdots \wedge x_n$.

Example 4.2. Let P be the set of finite (possibly empty) strings over letters a, b, c, and we define the order relation for $w, v \in P$ in a way that $w \preceq v$ if and only if w appears as a consecutive substring in v. For example, $aba \preceq cabab$ and $abb \not\preceq acbb$. Now for $w = ab$ and $v = bc$, the common lower bounds for w and v are the empty string (denoted here by $\emptyset$) and b; as $\emptyset \preceq b$, the meet of w and v is $w \wedge v = b$. On the other hand, the set of common upper bounds for w and v does not have the minimum element, therefore the join $w \vee v$ of w and v does not exist.

Definition 4.3. Let P be a poset. We say that P is a *lattice* if any finite non-empty subset of P has the join and the meet. We say that P is a *complete lattice* if any non-empty subset of P has the join and the meet.

We note that by definition, any finite lattice is a complete lattice.

Example 4.3.

- Any totally ordered set is a lattice.
- The set, say P, of the subgroups of a group G ordered by inclusion relation (i.e., $H_1 \preceq H_2 \Leftrightarrow H_1 \subseteq H_2$) is a complete lattice. Here the meet of all $H_\lambda \in P$ ($\lambda \in \Lambda$) is $\bigcap_{\lambda \in \Lambda} H_\lambda$, and their join is the subgroup of G generated by $\bigcup_{\lambda \in \Lambda} H_\lambda$.
- The set of positive integers $\mathbb{Z}_{>0}$ ordered by divisibility relation forms a lattice. Here the join of $a_1, \ldots, a_n \in \mathbb{Z}_{>0}$ is their least common multiple, and their meet is their greatest common divisor. On the other hand,

any non-empty subset of $\mathbb{Z}_{>0}$ has the meet (see the exercises), while $\mathbb{Z}_{>0}$ is not a complete lattice. Indeed, the subset $S = \{p \in \mathbb{Z}_{>0} \mid p \text{ is prime}\}$ does not have the join.

Proposition 4.1. *Let L be a non-empty complete lattice. Then L has the maximum element and the minimum element.*

Proof. As L is a complete lattice and $L \neq \emptyset$, the L itself has its join and meet. The former is the maximum element of L, and the latter is the minimum element of L. □

Proposition 4.2. *Let P be a poset satisfying the condition "any $x, y \in P$ have their join $x \vee y$". Then any finite non-empty subset of P has the join. An analogous property also holds when switching the large/small relations (concerning the meet instead of the join). In particular, if any two elements of P have their join and meet, then P is a lattice.*

Proof. It suffices to prove the claim for the join. For the case of singletons, we have obviously $\bigvee\{x\} = x$. From now, we prove that for any $x \in P$ and any finite non-empty subset $S \subseteq P$, the join $\bigvee(S \cup \{x\})$ exists and is equal to $(\bigvee S) \vee x$, by mathematical induction with respect to $|S|$. The case of S being a singleton $\{y\}$ is obvious by the hypothesis of the proposition. From now, we suppose that $|S| \geq 2$. The join $\bigvee S$ exists by the induction hypothesis. Now $a := (\bigvee S) \vee x$ is an upper bound of x, and it is an upper bound of an upper bound $\bigvee S$ of S, hence it is an upper bound of S as well. Therefore, a is an upper bound of $S \cup \{x\}$. On the other hand, for any upper bound b of $S \cup \{x\}$, we have $b \succeq x$, and as b is an upper bound of S, it follows from the definition of $\bigvee S$ that $b \succeq \bigvee S$. Hence we have $b \succeq (\bigvee S) \vee x = a$. Therefore a is the minimum upper bound of $S \cup \{x\}$, meaning that a is the join of $S \cup \{x\}$. Hence the current claim holds for any S. This implies the original claim. □

We explain a more efficient computation of the Möbius function μ_P of a poset P when P is a lattice. For a finite lattice L, let $A(L)$ denote the

$\mathbb{C}$-linear space with basis L, and define the multiplication over $A(L)$ as $\mathbb{C}$-algebra by $xy := x \wedge y$ $(x, y \in L)$. The multiplicative identity element of $A(L)$ is the maximum element 1_L of L. For any $x \in L$, we define

$$\zeta_x := \sum_{y \preceq x} \mu_L(y, x) y \in A(L) \ .$$

Lemma 4.1. *In this setting, for any $x \in L$ we have $x = \sum_{y \preceq x} \zeta_y$.*

Proof. For any $z \in L$, by writing as π_z the projection that maps each element of $A(L)$ to its coefficient of z, we have $\pi_z(\zeta_x) = \sum_{y \preceq x} \mu_L(y, x) \pi_z(y)$. Let $f_z, g_z \colon L \to \mathbb{C}$ with $f_z(x) := \pi_z(\zeta_x)$ and $g_z(x) := \pi_z(x)$. Then the equality above means the relation $f_z = g_z * \mu_L$ in $\mathbb{C}^L$. Now Möbius inversion formula implies that $g_z = f_z * \mathbb{1}$, therefore

$$\pi_z(x) = g_z(x) = \sum_{y \preceq x} \mathbb{1}(y, x) f_z(y) = \sum_{y \preceq x} \pi_z(\zeta_y) \ .$$

Hence we have $x = \sum_{y \preceq x} \zeta_y$ and the claim holds. □

By Lemma 4.1, the ζ_x's form a generating set of $A(L)$ as $\mathbb{C}$-linear space. Moreover, by comparing the numbers of elements, it also follows that the ζ_x's form a basis of $A(L)$ as $\mathbb{C}$-linear space.

Lemma 4.2. *If we define a $\mathbb{C}$-linear map θ from $A(L)$ to the direct sum of $|L|$ copies of $\mathbb{C}$ (where the canonical basis elements of the latter are denoted by $\widehat{x}$, $x \in L$) by $\theta(\zeta_x) := \widehat{x}$ $(x \in L)$, then θ is an isomorphism of $\mathbb{C}$-algebras. In particular, for $x, y \in L$, we have*

$$\zeta_x \zeta_y = \begin{cases} \zeta_x & (\text{if } x = y) \\ 0 & (\text{if } x \neq y). \end{cases}$$

Proof. The latter part of the claim is deduced from the former part and the definition of θ. It is obvious by definition that θ is an isomorphism of $\mathbb{C}$-linear spaces, therefore it suffices to show that θ preserves the multiplication.

For the purpose, it also suffices to show that for any $x, y \in L$ we have $\theta(x)\theta(y) = \theta(x \wedge y)$. Now we have

$$\begin{aligned}\theta(x)\theta(y) &= \theta\left(\sum_{z \preceq x} \zeta_z\right) \theta\left(\sum_{w \preceq y} \zeta_w\right) \\ &= \left(\sum_{z \preceq x} \widehat{z}\right) \left(\sum_{w \preceq y} \widehat{w}\right) \\ &= \sum_{z \preceq x\,,\, w \preceq y} \widehat{z}\widehat{w} \\ &= \sum_{z\,;\, z \preceq x\,,\, z \preceq y} \widehat{z} \\ &= \sum_{z \preceq x \wedge y} \widehat{z} = \theta\left(\sum_{z \preceq x \wedge y} \zeta_z\right) = \theta(x \wedge y) \ ,\end{aligned}$$

therefore the claim holds. □

For a poset P with maximum element 1_P, we say that an element of P is a *coatom* of P if it is covered by 1_P. Similarly, for a poset P with minimum element 0_P, we say that an element of P is an *atom* of P if it covers 0_P.

Theorem 4.1. *Let L be a finite lattice, and let A^* denote the set of the coatoms of L. For any non-negative integer k, put*

$$N_k := \left| \left\{ S \subseteq A^* \colon |S| = k \,,\, \bigwedge S = 0_L \right\} \right| \ .$$

Then we have $\mu_L(0_L, 1_L) = \sum_{k \geq 0} (-1)^k N_k$.

Proof. For any $x \in A^*$, the following holds in $A(L)$:

$$1_L - x = \sum_{y \preceq 1_L} \zeta_y - \sum_{y \preceq x} \zeta_y = \sum_{y\,,\, y \not\preceq x} \zeta_y \ .$$

Therefore we have

$$\prod_{x \in A^*} (1_L - x) = \prod_{x \in A^*} \sum_{y \not\preceq x} \zeta_y \ .$$

When expanding the right-hand side into a linear combination of the ζ_y's by using Lemma 4.2, the terms not vanishing are the terms of ζ_y's satisfying that for any $x \in A^*$ we have $y \not\preceq x$. By the definition of A^*, the only such y is $y = 1_L$; therefore we have

$$\prod_{x \in A^*} (1_L - x) = \zeta_{1_L} = \sum_{z \in L} \mu_L(z, 1_L) z \ .$$

By comparing the coefficients of 0_L in both sides, it follows that

$$\sum_{k \geq 0} (-1)^k N_k = \mu_L(0_L, 1_L) \ .$$

Hence the claim holds. □

Several subclasses of lattices satisfying certain additional conditions have been studied. Here we explain an example among them.

Definition 4.4. We say that a lattice L is a *distributive lattice* if the distributive law holds; that is, for any $x, y, z \in L$ we have $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ and $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$.

Remark 4.1. We note that for any elements x, y, z of a lattice L, we always have $x \vee (y \wedge z) \preceq (x \vee y) \wedge (x \vee z)$ and $x \wedge (y \vee z) \succeq (x \wedge y) \vee (x \wedge z)$. Therefore, among the conditions for distributive lattices, it suffices to verify that $x \vee (y \wedge z) \succeq (x \vee y) \wedge (x \vee z)$ and $x \wedge (y \vee z) \preceq (x \wedge y) \vee (x \wedge z)$.

Example 4.4.

- The power set $\mathcal{P}(S)$ of a given set S, ordered by the inclusion relation, forms a distributive lattice. Indeed, the join and the meet in $\mathcal{P}(S)$ correspond to the set union and the set intersection, respectively, and the condition for a distributive lattice is now nothing but the distributive law for set union and set intersection.

- Let P be a poset. For a subset $I \subseteq P$, we say that I is an *order ideal* of P if for any $x, y \in P$, the conditions $x \in I$ and $y \preceq x$ imply that $y \in I$. We define

$$J(P) := \{I \subseteq P \mid I \text{ is an order ideal of } P\} \ .$$

Then $(J(P), \subseteq)$ is a distributive lattice. Indeed, for any $I_1, I_2 \in J(P)$, we have $I_1 \cap I_2 \in J(P)$ and $I_1 \cup I_2 \in J(P)$, therefore $I_1 \wedge I_2 = I_1 \cap I_2$ and $I_1 \vee I_2 = I_1 \cup I_2$; hence $J(P)$ is a lattice. Now the condition for a distributive lattice is deduced from the properties of set inclusion.

- The lattice, say L, consisting of the subgroups of a given group G is in general not a distributive lattice. To see this, we consider $G = S_3$ and its subgroups $H_1 = \langle(12)\rangle$, $H_2 = \langle(13)\rangle$, and $H_3 = \langle(23)\rangle$. Now we have $H_2 \wedge H_3 = \{\mathsf{id}\}$ and hence $H_1 \vee (H_2 \wedge H_3) = H_1 \vee \{\mathsf{id}\} = H_1$. On the other hand, we have $H_1 \vee H_2 = H_1 \vee H_3 = S_3$ and hence $(H_1 \vee H_2) \wedge (H_1 \vee H_3) = S_3 \wedge S_3 = S_3$. This means that $H_1 \vee (H_2 \wedge H_3) \neq (H_1 \vee H_2) \wedge (H_1 \vee H_3)$, therefore L is not distributive.

The following characterization is known for distributive lattices.

Theorem 4.2. *Let L be a finite distributive lattice. Then there exists a unique poset P, up to isomorphism, satisfying $L \simeq J(P)$.*

Proof. First, we show the existence of such a P. We say that an element x of L is *join-irreducible* if there exist no $y, z \in L$ with $y, z \prec x$ and $y \vee z = x$. Put

$$P := \{x \in L \mid x \text{ is join-irreducible}\}$$

and define maps $f\colon L \to J(P)$ and $g\colon J(P) \to L$ by

$$f(x) := P \cap \wedge^x \,,\ g(I) := \bigvee I \ ,$$

where we set $\wedge^x := \{y \in L \mid y \preceq x\}$. For any $x, y \in L$ with $x \preceq y$, we have $\wedge^x \subseteq \wedge^y$, therefore $f(x) \subseteq f(y)$. This means that f is order-preserving. Similarly, for any $I_1, I_2 \in J(P)$ with $I_1 \subseteq I_2$, we have $\bigvee I_1 \preceq \bigvee I_2$, therefore $g(I_1) \preceq g(I_2)$. This means that g is order-preserving as well.

We show that $g(f(x)) = x$ for any $x \in L$, by using mathematical induction with respect to the maximum number of edges involved in an upward path from 0_L to x in a Hasse diagram of L, denoted by $\rho(x)$. When $x \in P$

(including the case of $\rho(x) = 0$ or equivalently $x = 0_L$), as x is the maximum element of $f(x)$ by the definition of f, it follows that $g(f(x)) = x$. From now, we consider the other case where $x \notin P$. By the definition of P, there are $y, z \in L$ with $y, z \prec x$ and $y \vee z = x$. Then we have $\rho(x) > \rho(y)$ and $\rho(x) > \rho(z)$, therefore the induction hypothesis implies that $g(f(y)) = y$ or equivalently $y = \bigvee(P \cap \wedge^y)$, and that $g(f(z)) = z$ or equivalently $z = \bigvee(P \cap \wedge^z)$. Now as $y \prec x$, we have $P \cap \wedge^y \subseteq P \cap \wedge^x$, therefore $\bigvee(P \cap \wedge^y) \preceq \bigvee(P \cap \wedge^x)$; and similarly, we also have $\bigvee(P \cap \wedge^z) \preceq \bigvee(P \cap \wedge^x)$. Moreover, as x is an upper bound of $P \cap \wedge^x$, we have $\bigvee(P \cap \wedge^x) \preceq x$. These arguments imply

$$x = y \vee z = \left(\bigvee(P \cap \wedge^y)\right) \vee \left(\bigvee(P \cap \wedge^z)\right) \preceq \bigvee(P \cap \wedge^x) \preceq x \ ,$$

therefore we have $x = \bigvee(P \cap \wedge^x) = g(f(x))$. Hence we have $g(f(x)) = x$ for any $x \in L$.

We show, conversely, that $f(g(I)) = I$ for any $I \in J(P)$. When $x \in I$ $(\subseteq P)$, we have $x \preceq g(I)$ and hence $x \in \wedge^{g(I)}$, therefore $x \in f(g(I))$. Hence we have $I \subseteq f(g(I))$. Conversely, when $x \in f(g(I))$, the definitions of f and g imply that $x \preceq \bigvee I$. As L is a distributive lattice, we have $x = x \wedge (\bigvee I) = \bigvee_{y \in I}(x \wedge y)$. Now the fact $x \in P$ implies that x is join-irreducible, therefore it follows that there exists a $y \in I$ satisfying that $x \wedge y = x$, hence $x \preceq y$. As I is an order ideal, it follows that $x \in I$. Hence we have $f(g(I)) \subseteq I$. Summarizing, we have $f(g(I)) = I$. By these results, f and g are the inverses of each other, therefore f is an isomorphism from L to $J(P)$. Hence the existence of a P as in the statement is proved.

From now, we show the uniqueness of such a P. For the purpose, we show that for any finite poset Q and any $I \in J(Q)$, I is join-irreducible in $J(Q)$ if and only if $I = \wedge^x$ for a unique $x \in Q$. When the latter condition holds, for any $K_1, K_2 \in J(Q)$ with $K_1 \cup K_2 = I$, as $x \in I$, we have $x \in K_i$ for at least one i. Now as K_i is an order ideal, we have $I = \wedge^x \subseteq K_i$; while the fact $K_1 \cup K_2 = I$ implies that $I \supseteq K_i$. Hence we have $K_i = I$. This implies that I is join-irreducible.

Conversely, suppose that the former condition holds. As I is a finite order ideal, we have $I = \bigcup_{y \in I\,;\,\mathrm{maximal}} \wedge^y$. For each such y, we have $\wedge^y \in J(Q)$. Now as I is join-irreducible, it follows that we have $I = \wedge^y$ for some maximal element y of I. On the other hand, if some $z \in Q$ also satisfies that $I = \wedge^z$, then we have $y \in I = \wedge^z$ and hence $y \preceq z$, and similarly we also have $z \preceq y$; therefore $z = y$. Hence the latter condition holds. Therefore the equivalence mentioned above is proved. Consequently, we can construct a map from the set of join-irreducible elements of $J(Q)$, denoted by $L(J(Q))$, to Q by associating to $I = \wedge^x$ the element $x \in Q$; this is an isomorphism from $L(J(Q))$ to Q.

Now suppose that for posets P_1 and P_2 we have $L \simeq J(P_1) \simeq J(P_2)$. Then the $J(P_i)$'s are finite, and the $\wedge^x$'s with $x \in P_i$ are distinct elements of $J(P_i)$. Hence we have $|P_i| < \infty$. Now as the join-irreducible property is preserved by isomorphisms, the aforementioned isomorphism $L(J(P_i)) \simeq P_i$ implies that $P_1 \simeq L(J(P_1)) \simeq L(J(P_2)) \simeq P_2$. Hence the P as in the statement is unique up to isomorphism. This completes the proof. □

Exercises

Problem 1. Let P be a poset.

1. Prove that if $x \in P$ is a maximum element of P, then x is a unique maximal element of P.

2. Suppose that P is finite. Prove that if $x \in P$ is a unique maximal element of P, then x is a maximum element of P.

3. When P is infinite, the property "if $x \in P$ is a unique maximal element of P, then x is a maximum element of P" does not hold in general. Prove this fact by giving an example of such a P.

(Comment: The analogous properties also hold when switching the sides of the order relation.)

Problem 2. In the setting of Example 4.2, prove that the elements $w = ab$ and $v = bc$ of P do not have the join $w \vee v$.

Problem 3. Suppose that a poset P is locally finite, has the minimum element 0_P, and satisfies that any finite non-empty subset of P has the join. Prove that any non-empty subset of P has the join.

Problem 4. Prove that for any elements x, y, z of a lattice L, we have $x \vee (y \wedge z) \preceq (x \vee y) \wedge (x \vee z)$.

(Comment: We can also prove that $x \wedge (y \vee z) \succeq (x \wedge y) \vee (x \wedge z)$ similarly.)

5 Well-Order and Mathematical Induction

This section is based on Chapter 1 of the book [4].

In the principle of mathematical induction, the properties of large/small relations on the set of non-negative integers are significantly relevant. By abstracting the properties, we are led to the following definition.

Definition 5.1. For a poset $(P, \preceq)$, we say that P is a *well-ordered set* and $\preceq$ is a *well-order* if any non-empty subset of P has the minimum element. When P is a well-ordered set, for any $x \in P$, we define $\mathrm{prec}_P(x) := \{y \in P \mid y \prec x\}$ and call it the *initial segment* of P by x.

Lemma 5.1. *Any well-order is a total order.*

Proof. For any well-ordered set P and $x, y \in P$, by definition, $\{x, y\}$ has the minimum element, say z. When $z = x$, the definition of z implies that $x \preceq y$. Similarly, when $z = y$ we have $y \preceq x$. Hence we have either $x \preceq y$ or $y \preceq x$, therefore P is a totally ordered set, as desired. □

The following property can be seen as "mathematical induction on well-ordered sets".

Theorem 5.1. *Let P be a non-empty well-ordered set, and let φ be some proposition for an element of P. If both "φ is true for $\min P$" and "for any $x \in P \setminus \{\min P\}$, if φ is always true on $\mathrm{prec}_P(x)$, then φ is also true for x" hold, then φ is always true on P.*

Proof. It suffices to deduce a contradiction by assuming that φ is false for some element of P. By this assumption, $\{x \in P \mid \varphi(x) \text{ is false}\}$ is a non-empty subset of P, therefore it has the minimum element, say a. Now the former condition in the statement implies that $a \neq \min P$, and the definition of a implies that φ is always true on $\mathrm{prec}_P(a)$. Therefore, the latter condition in the statement implies that φ is true for a as well, a contradiction. Hence the claim holds. □

The set $\mathbb{Z}_{\geq 0}$ of non-negative integers is a well-ordered set. By applying Theorem 5.1 to the $\mathbb{Z}_{\geq 0}$, we can deduce the usual mathematical induction (or its variant, so-called course-of-values induction). Moreover, the fact that $\mathbb{Z}_{\geq 0}$ is a well-ordered set also implies the following property.

Theorem 5.2 (Pigeonhole Principle)**.** *Let $n > m \geq 1$ be integers. Then any map $f\colon \{1, \ldots, n\} \to \{1, \ldots, m\}$ is not injective.*

Proof. Assume for the contrary that some m does not satisfy the claim. Take the smallest such m, and suppose that a map f in the statement is injective. It holds obviously that $m > 1$. As f is injective, $|f^{-1}[m]|$ is either 0 or 1. If $|f^{-1}[m]| = 0$, then $f\colon \{1, \ldots, n\} \to \{1, \ldots, m-1\}$ is also injective, contradicting the minimality of m. Hence we must have $|f^{-1}[m]| = 1$. By writing the unique element of $f^{-1}[m]$ as a, there exists a bijection $g\colon \{1, \ldots, n-1\} \to \{1, \ldots, n\} \setminus \{a\}$. Now $f \circ g\colon \{1, \ldots, n-1\} \to \{1, \ldots, m-1\}$ is also injective, contradicting the minimality of m. Hence the claim holds. □

In order to show important properties of well-orders, we prepare some properties.

Lemma 5.2. *For any well-ordered sets $(P, \preceq_P)$ and $(Q, \preceq_Q)$, any isomorphism $f\colon P \to Q$, and any $x \in P$, we have $f(\mathrm{prec}_P(x)) = \mathrm{prec}_Q(f(x))$.*

Proof. As f is order-preserving, we have $f(\mathrm{prec}_P(x)) \subseteq \mathrm{prec}_Q(f(x))$. Similarly, as f^{-1} is order-preserving, we have $f^{-1}(\mathrm{prec}_Q(f(x))) \subseteq \mathrm{prec}_P(x)$ and hence $\mathrm{prec}_Q(f(x)) \subseteq f(\mathrm{prec}_P(x))$. Therefore we have $f(\mathrm{prec}_P(x)) = \mathrm{prec}_Q(f(x))$ and the claim holds. □

Lemma 5.3. *For any well-ordered set $(P, \preceq)$ and any $x \in P$, we have $P \ncong \mathrm{prec}_P(x)$.*

Proof. Assume for the contrary that an isomorphism $f\colon \mathrm{prec}_P(x) \to P$ exists. To show that $S := \{y \in \mathrm{prec}_P(x) \mid f(y) \neq y\}$ is empty, assume for the

contrary that S is non-empty. Take the minimum element y of S. Then $f|_{\mathrm{prec}_P(y)}$ has to be an identity map, and the restriction of f to $\mathrm{prec}_P(x) \setminus \mathrm{prec}_P(y)$ is an isomorphism onto $P \setminus \mathrm{prec}_P(y)$. This isomorphism has to map the minimum element y of $\mathrm{prec}_P(x) \setminus \mathrm{prec}_P(y)$ to the minimum element y of $P \setminus \mathrm{prec}_P(y)$, therefore $f(y) = y$. This contradicts the fact $y \in S$; therefore we have $S = \emptyset$. This implies that for any $y \in \mathrm{prec}_P(x)$ we have $f(y) = y$, therefore there is no $y \in \mathrm{prec}_P(x)$ with $f(y) = x \in P$, contradicting the fact that f is surjective. Hence the claim holds. □

Lemma 5.4. *For any well-ordered sets $(P, \preceq_P)$ and $(Q, \preceq_Q)$, an isomorphism from P to Q is unique if it exists.*

Proof. Let $f, g\colon P \to Q$ be isomorphisms. Assume for the contrary that there is some $x \in P$ with $f(x) \neq g(x)$. For the minimum such x, we have $f|_{\mathrm{prec}_P(x)} = g|_{\mathrm{prec}_P(x)}$. By writing its image as X, both f and g are isomorphisms from $P \setminus \mathrm{prec}_P(x)$ onto $Q \setminus X$. As x is the minimum element of $P \setminus \mathrm{prec}_P(x)$, both $f(x)$ and $g(x)$ have to be the minimum element of $Q \setminus X$, contradicting the choice of x. Hence we have $f = g$ and the claim holds. □

The following property is remarkable for well-ordered sets.

Theorem 5.3. *Let $(P, \preceq_P)$ and $(Q, \preceq_Q)$ be well-ordered sets. Then precisely one of the following conditions holds:*

1. *$P \simeq Q$.*
2. *There exists an $x \in P$ with $\mathrm{prec}_P(x) \simeq Q$.*
3. *There exists an $y \in Q$ with $P \simeq \mathrm{prec}_Q(y)$.*

Proof. First, we show that any two conditions in the claim do not hold at once. Suppose that Condition 1 holds. If Condition 2 also holds, then we have $P \simeq \mathrm{prec}_P(x)$, contradicting Lemma 5.3. The case where Condition 3 instead holds is similarly contradictory. Hence Condition 1 cannot be consistent

with the other conditions. On the other hand, assume for the contrary that Conditions 2 and 3 hold. By choosing an isomorphism $f\colon P \to \mathrm{prec}_Q(y)$, Lemma 5.2 implies that $f(\mathrm{prec}_P(x)) = \mathrm{prec}_{\mathrm{prec}_Q(y)}(f(x)) = \mathrm{prec}_Q(f(x))$. Now the composition of it and the isomorphism in Condition 2 yields an isomorphism $Q \to \mathrm{prec}_Q(f(x))$, contradicting Lemma 5.3. Hence any two conditions in the claim do not hold at once.

Our remaining task is to show that at least one of the conditions in the claim holds. Let $\mathcal{F}$ denote the set of isomorphisms from a subset $\mathrm{prec}_P(a) \cup \{a\} \subseteq P$ with some $a \in P$ to a subset $\mathrm{prec}_Q(b) \cup \{b\} \subseteq Q$ with some $b \in Q$. Let $X \subseteq P$ be the union of the domains of maps in $\mathcal{F}$. We are going to define a map $F\colon X \to Q$ in a way that if $a \in X$ belongs to the domain of $f \in \mathcal{F}$ then $F(a) := f(a)$. Assume for the contrary that there is some $a \in X$ that belongs to the domains of both $f, g \in \mathcal{F}$ and satisfies that $f(a) \neq g(a)$. Take the minimum such a and the corresponding f, g. By the definition of $\mathcal{F}$, the domain of any of f and g includes $\mathrm{prec}_P(a) \cup \{a\}$. By Lemma 5.2 we have $f(\mathrm{prec}_P(a)) = \mathrm{prec}_Q(f(a))$ and $g(\mathrm{prec}_P(a)) = \mathrm{prec}_Q(g(a))$, while the definition of a implies that f and g coincide on $\mathrm{prec}_P(a)$; therefore we have $\mathrm{prec}_Q(f(a)) = \mathrm{prec}_Q(g(a))$ and hence $f(a) = g(a)$. This contradicts the choice of a. Therefore, if $a \in X$ belongs to the domains of both $f, g \in \mathcal{F}$, then we have $f(a) = g(a)$. This means that such a map F is well-defined.

We note that by the definition of $\mathcal{F}$, for any $x, y \in X$, there is an $f \in \mathcal{F}$ whose domain involves both x, y (indeed, it suffices to take any $f \in \mathcal{F}$ whose domain involves the maximum among x and y). For any $x, y \in X$ with $x \prec_P y$, by taking an $f \in \mathcal{F}$ whose domain involves x, y, the fact that f is an isomorphism implies that $F(x) = f(x) \prec_Q f(y) = F(y)$. Hence F is order-preserving and injective. Put $Y := F(X)$. For any $x, y \in X$ with $F(x) \prec_Q F(y)$, by taking an $f \in \mathcal{F}$ whose domain involves x, y, we have $f(x) = F(x) \prec_Q F(y) = f(y)$; as f is an isomorphism, we have $x \prec_P y$. Hence F is an isomorphism from X onto Y.

As both the domain and the range of an element of $\mathcal{F}$ are order ideals,

the sets X and Y being their unions, respectively, are order ideals as well. Therefore, if $X \neq P$ then the element $x := \min(P \setminus X)$ satisfies that $X = \mathrm{prec}_P(x)$, and similarly, if $Y \neq Q$ then the element $y := \min(Q \setminus Y)$ satisfies that $Y = \mathrm{prec}_Q(y)$. This implies that some of the conditions in the claim holds when $X = P$ or $Y = Q$. Hence our task is reduced to deduce a contradiction by assuming that $X \neq P$ and $Y \neq Q$. By expressing as $X = \mathrm{prec}_P(x)$ and $Y = \mathrm{prec}_Q(y)$ as above, we can extend the map F to an isomorphism $\overline{F}\colon X \cup \{x\} \to Y \cup \{y\}$ by defining $\overline{F}(x) := y$. Now we have $\overline{F} \in \mathcal{F}$ by definition, contradicting the fact $x \notin X$. This completes the proof. □

The usual mathematical induction consists of "n-th steps" for *finite* numbers n; we want to extend it to handle the cases of *infinite* numbers n as well. For the purpose, we introduce the notion of ordinal numbers. In the following argument, when we are focusing on ordinal numbers, we suppose that every element of any set is also a set. That is, for any two elements x, y of some set, we can deduce that $x = y$ by showing "$a \in x$ if and only if $a \in y$".

Definition 5.2. We say that a set x is *transitive* if for any $y \in x$ and any $z \in y$ we have $z \in x$ (or equivalently, for any $y \in x$ we have $y \subseteq x$). We say that a set α is an *ordinal number* if the following conditions hold:

1. α is transitive.
2. If $\beta \in \alpha$, then we have $\beta \notin \beta$.
3. The relation $\preceq_\alpha$ on α, defined in a way that $\beta \preceq_\alpha \gamma$ if and only if either $\beta = \gamma$ or $\beta \in \gamma$, is a well-order.

We write the collection (class) of all ordinal numbers as ON.

It is known that ON itself is not a set (if we treat ON as a set, then a contradiction occurs). This fact is named *Burali-Forti paradox*. We note also that by Condition 2 above, for any ordinal number α we have $\alpha \notin \alpha$.

Example 5.1. By defining $\ulcorner n \urcorner$ for each non-negative integer n in such a way that $\ulcorner 0 \urcorner := \emptyset$, $\ulcorner 1 \urcorner := \{\ulcorner 0 \urcorner\}$, $\ulcorner 2 \urcorner := \{\ulcorner 0 \urcorner, \ulcorner 1 \urcorner\}$, $\ulcorner 3 \urcorner := \{\ulcorner 0 \urcorner, \ulcorner 1 \urcorner, \ulcorner 2 \urcorner\}, \dots$, it follows that they are ordinal numbers. We call them *finite ordinal numbers*. (From now on, we identify a non-negative integer n with $\ulcorner n \urcorner$.) On the other hand, $\omega := \{\ulcorner n \urcorner \mid n \in \mathbb{Z}_{\geq 0}\}$ and $\omega \cup \{\omega\}$ are also ordinal numbers. They are examples of infinite ordinal numbers.

From now on, we explain several properties of ordinal numbers.

Proposition 5.1. *If α is an ordinal number and $\beta \in \alpha$, then β is also an ordinal number and we have* $\mathrm{prec}_\alpha(\beta) = \beta$.

Proof. For the latter part of the claim, if $\gamma \in \mathrm{prec}_\alpha(\beta)$ then we have $\gamma \prec_\alpha \beta$, therefore $\gamma \in \beta$. Hence we have $\mathrm{prec}_\alpha(\beta) \subseteq \beta$. Conversely, for any $\gamma \in \beta$, the hypothesis $\beta \in \alpha$ and the transitivity of α imply that $\gamma \in \alpha$; as $\gamma \in \beta$, we have $\gamma \prec_\alpha \beta$, therefore $\gamma \in \mathrm{prec}_\alpha(\beta)$. Hence we have $\beta \subseteq \mathrm{prec}_\alpha(\beta)$, therefore $\mathrm{prec}_\alpha(\beta) = \beta$.

For the former part of the claim, the latter part of the claim implies that β is a subset of α; hence β is also a well-ordered set, and the property "$\gamma \in \beta$ implies $\gamma \notin \gamma$" is inherited from α. On the other hand, suppose that $\delta \in \gamma \in \beta$. Then the hypothesis $\beta \in \alpha$ and the transitivity of α imply that $\gamma \in \alpha$ and moreover that $\delta \in \alpha$. Now we have $\delta \prec_\alpha \gamma \prec_\alpha \beta$ and hence $\delta \prec_\alpha \beta$, therefore $\delta \in \beta$. This implies that β is also transitive, therefore β is also an ordinal number. Hence the claim holds. □

Proposition 5.2. *For any ordinal numbers α, β, if α and β are isomorphic as well-ordered sets, then $\alpha = \beta$.*

Proof. Take an isomorphism $f \colon \alpha \to \beta$. Assume for the contrary that $S := \{\gamma \in \alpha \mid f(\gamma) \neq \gamma\}$ is non-empty. Let γ be the minimum element of S.

For any $\delta \in \gamma$, the fact $\gamma \in \alpha$ and the transitivity of α imply that $\delta \in \alpha$. Therefore we have $\delta \prec_\alpha \gamma$, and by the definition of γ we have $f(\delta) = \delta$. As f is order-preserving, the fact $\delta \in \gamma$ implies that $\delta = f(\delta) \in f(\gamma)$. Hence we have $\gamma \subseteq f(\gamma)$.

For any $\delta \in f(\gamma)$, the fact $f(\gamma) \in \beta$ and the transitivity of β imply that $\delta \in \beta$. By putting $\eta := f^{-1}(\delta) \in \alpha$, we have $f(\eta) = \delta \in f(\gamma)$; as f is an isomorphism, we have $\eta \in \gamma$, therefore $\eta \in \mathrm{prec}_\alpha(\gamma)$. Hence by the definition of γ, we have $f(\eta) = \eta$ and $\delta = f(\eta) = \eta \in \gamma$. Hence we have $f(\gamma) \subseteq \gamma$, therefore $f(\gamma) = \gamma$. This contradicts the fact $\gamma \in S$.

The argument above implies that $S = \emptyset$, that is, for any $\gamma \in \alpha$ we have $f(\gamma) = \gamma$. Now for any $\gamma \in \alpha$, we have $\gamma = f(\gamma) \in \beta$. Therefore we have $\alpha \subseteq \beta$. On the other hand, for any $\gamma \in \beta$, we have $\gamma = f(f^{-1}(\gamma)) = f^{-1}(\gamma) \in \alpha$. Therefore we have $\beta \subseteq \alpha$. Hence we have $\alpha = \beta$ and the claim holds. □

Theorem 5.4. *For any ordinal numbers α, β, precisely one of $\alpha = \beta$, $\alpha \in \beta$, and $\beta \in \alpha$ holds.*

Proof. First, as $\alpha \notin \alpha$, if $\alpha = \beta$ then we have $\alpha \notin \beta$ and $\beta \notin \alpha$. On the other hand, if both $\alpha \in \beta$ and $\beta \in \alpha$ hold, then the transitivity of α implies that $\alpha \in \alpha$, contradicting the fact above. Hence two conditions in the claim do not hold at once.

By Theorem 5.3, one of the following conditions holds: $\alpha \simeq \beta$; there is a $\gamma \in \alpha$ with $\mathrm{prec}_\alpha(\gamma) \simeq \beta$; and there is a $\gamma \in \beta$ with $\alpha \simeq \mathrm{prec}_\beta(\gamma)$. In the first case, Proposition 5.2 implies that $\alpha = \beta$. In the second case, Proposition 5.1 implies that γ is an ordinal number and $\mathrm{prec}_\alpha(\gamma) = \gamma$. Therefore Proposition 5.2 implies that $\gamma = \beta$ and $\beta \in \alpha$. In the third case, we similarly have $\alpha \in \beta$. Hence the claim holds. □

Theorem 5.5. *Let S be a set consisting of ordinal numbers, and for $\alpha, \beta \in S$, we define $\alpha \leq \beta$ if and only if either $\alpha = \beta$ or $\alpha \in \beta$. Then $(S, \leq)$ is a well-ordered set.*

Proof. First, we show that $(S, \leq)$ is a poset. The reflexivity follows immediately by definition. For the antisymmetry, it suffices to deduce a contradiction by assuming that $\alpha \leq \beta$, $\beta \leq \alpha$, and $\alpha \neq \beta$. Now the definition of $\leq$ implies that $\alpha \in \beta$ and $\beta \in \alpha$, contradicting Theorem 5.4, as desired. For the transitivity, we suppose that $\alpha \leq \beta$ and $\beta \leq \gamma$ and show that $\alpha \leq \gamma$. This is

obvious when $\alpha = \beta$ or $\beta = \gamma$; we consider the other case where $\alpha \neq \beta$ and $\beta \neq \gamma$. Now the definition of $\leq$ implies that $\alpha \in \beta$ and $\beta \in \gamma$, therefore the transitivity of γ implies that $\alpha \in \gamma$ and $\alpha \leq \gamma$, as desired. Hence $(S, \leq)$ is a poset.

The remaining task is to show that if $\emptyset \neq T \subseteq S$, then T has the minimum element with respect to $\leq$. Take an element $\alpha \in T$. If the α is the minimum element of T then the claim holds; we suppose that we are not in this case. Now $X := \{\beta \in T \mid \alpha \not\leq \beta\}$ is non-empty. Moreover, for any $\beta \in X$, we have $\alpha \neq \beta$ and $\alpha \not\in \beta$, therefore by Theorem 5.4 we have $\beta \in \alpha$. Hence we have $X \subseteq \alpha$. As α is a well-ordered set, X has the minimum element, say γ. If both $\delta \in T$ and $\gamma \not\leq \delta$ were satisfied, then as γ is the minimum element of X, we would have $\delta \not\in X$ and hence $\alpha \leq \delta$. On the other hand, the fact $\gamma \in X \subseteq \alpha$ implies that $\gamma \leq \alpha$. Now the transitivity of $\leq$ would imply that $\gamma \leq \delta$, a contradiction. Hence for any $\delta \in T$ we have $\gamma \leq \delta$, therefore γ is the minimum element of T, as desired. Hence the claim holds. □

Remark 5.1. An argument similar to Theorem 5.5 implies that any non-empty collection of ordinal numbers involves its minimum element.

In the following argument, we suppose that the well-order $\leq$ as in Theorem 5.5 is defined for ordinal numbers.

By using ordinal numbers, we can formalize an argument like "mathematical induction of infinite length", formally called *transfinite induction* or *transfinite recursion*. Here we omit the precise formulation of transfinite induction as it is too technical; we instead explain a proof of Zorn's Lemma from the Axiom of Choice as an example of transfinite induction. For the purpose, we give the following definitions.

Definition 5.3. Let α be an ordinal number. We say that α is a *successor ordinal* if α has the maximum element as ordered set. We say that α is a *limit ordinal* if $\alpha \neq 0$ and α is not a successor ordinal.

We note that when α is a successor ordinal and $\beta = \max \alpha$, β is the maximum ordinal number less than α.

Theorem 5.6 (Zorn's Lemma)**.** *Let P be a non-empty poset. If any totally ordered subset of P has an upper bound in P, then P has a maximal element.*

Proof. First of all, by applying the Axiom of Choice to the family of all non-empty subsets of P, we can take a map ι with domain being the set of non-empty subsets of P, for which we always have $\iota(S) \in S$.

Assume for the contrary that P does not have a maximal element. For each ordinal number α, we recursively define an element x_α of P satisfying the condition "$\beta < \alpha$ implies $x_\beta \prec x_\alpha$" in the following manner. For $\alpha = 0$ we define $x_0 := \iota(P)$. We consider the case where $\alpha \neq 0$, and suppose that the x_β's have been defined for ordinal numbers β less than α.

- When α is a successor ordinal, let β denote the maximum element of α. By the assumption on P, $x_\beta \in P$ is not a maximal element, therefore $S := \{y \in P \mid x_\beta \prec y\}$ is non-empty. We define $x_\alpha := \iota(S)$. Now we have $x_\beta \prec x_\alpha$ by definition, while for any ordinal number γ less than β, the condition $x_\gamma \prec x_\beta$ implies that $x_\gamma \prec x_\alpha$. Hence the condition above is satisfied by the x_α.

- When α is a limit ordinal, by putting $C := \{x_\beta \mid \beta < \alpha\}$, the condition above implies that C is a totally ordered subset of P. By the hypothesis, $S := \{y \in P \mid y \text{ is an upper bound of } C\}$ is non-empty. We define $x_\alpha := \iota(S)$. Now let $\beta < \alpha$. Then we have $x_\beta \preceq x_\alpha$. Moreover, as α is a limit ordinal, β is not the maximum ordinal number less than α, implying that there is an ordinal number γ with $\beta < \gamma < \alpha$. Now we have $x_\beta \prec x_\gamma \preceq x_\alpha$, therefore $x_\beta \prec x_\alpha$. Hence the condition above is satisfied by the x_α.

By the argument, an element x_α satisfying the condition above is defined for every ordinal number α (intuitively, if there were an α for which x_α is not

defined, then we could take the minimum such α, but for this α the element x_α should be defined by the argument above, a contradiction). Owing to the condition above, those x_α's are all distinct. This implies (intuitively; we omit a formal argument here) that "the collection of all ordinal numbers corresponding to some element of the set P via the rule above" forms a set, which is nothing but the ON itself, contradicting the Burali-Forti paradox. Hence the claim holds. □

A standard argument using Zorn's Lemma implies the following property.

Theorem 5.7 (Well-Ordering Theorem)**.** *Any set becomes a well-ordered set with respect to some order relation.*

Based on Well-Ordering Theorem and the notion of ordinal numbers, we can give a (set-theoretic) definition of cardinalities of sets.

Theorem 5.8. *Let X be a set. Then there exist an ordinal number α and a bijection $X \to \alpha$. We call the minimum such ordinal number α the* cardinality *of X, denoted by $|X|$.*

Proof. By Well-Ordering Theorem, we may suppose without loss of generality that X is a well-ordered set. By Theorem 5.3, any ordinal number satisfies one of the conditions in Theorem 5.3 with X. Now for each $x \in X$, Proposition 5.2 implies that there is at most one ordinal number α with $\mathrm{prec}_X(x) \simeq \alpha$. Based on this fact, we (again intuitively) define S to be "the set of ordinal numbers corresponding to elements of X in the way above". Then, as ON itself is not a set by the Burali-Forti paradox, it follows that S is not equal to ON, therefore there is an ordinal number α not belonging to S. For this α, among the conditions in Theorem 5.3, either "$X \simeq \alpha$" or "there is a $\beta \in \alpha$ with $X \simeq \mathrm{prec}_\alpha(\beta) = \beta$" holds (where the last equality follows from Proposition 5.1). In any case, there is a bijection from X to an ordinal number α or β. Hence the claim holds. □

Exercises

Problem 1. Let $(P, \preceq_P)$ and $(Q, \preceq_Q)$ be well-ordered sets. We define a relation $\preceq$ on the disjoint union $P \sqcup Q$ of P and Q in a way that it coincides with $\preceq_P$ and $\preceq_Q$ on P and Q, respectively, and for any $x \in P$ and $y \in Q$ we have $x \preceq y$. Prove that this $\preceq$ is a well-order.

Problem 2. Let $(P, \preceq_P)$ and $(Q, \preceq_Q)$ be well-ordered sets. We define a relation $\preceq$ on the direct product $P \times Q$ of P and Q in a way that $(x_1, x_2) \preceq (y_1, y_2)$ if and only if either $x_2 \prec y_2$ or "$x_2 = y_2$ and $x_1 \preceq y_1$". Prove that this $\preceq$ is a well-order.

(Comment: Such an order is called a lexicographic order.)

Problem 3. Let α, β be ordinal numbers. Prove that $\alpha \in \beta$ if and only if $\alpha \subsetneq \beta$.

Problem 4. Prove the Burali-Forti paradox.

(Hint: Observe that if ON is a set, then ON itself becomes an ordinal number.)

Problem 5. Prove Well-Ordering Theorem (by using Zorn's Lemma).

6 Graph Theory

In this and the following sections, the descriptions about graph theory are based on the books [2] and [3].

Graphs are a tool of representing situations where certain objects are "joined/not joined" to each other.

Definition 6.1. An *undirected graph* is defined as a triplet (V, E, ι) of two sets V, E and a map $\iota\colon E \to \{\{x, y\} \mid x, y \in V\}$.

When expressing an undirected graph, we often omit the symbol ι and simply write (V, E). Moreover, we sometimes call an undirected graph simply a graph. When V and E are finite sets, we say that the undirected graph (V, E, ι) is finite.

For an undirected graph $G = (V, E, \iota)$, we call V the *vertex set* of G, and call E the *edge set* of G. Any element of V and E is called a *vertex* and an *edge* of G, respectively. When we want to emphasize that an edge of G is an edge of an undirected graph, we sometimes call it an undirected edge. Intuitively, when vertices x, y and an edge e of G satisfy that $\iota(e) = \{x, y\}$, this e can be regarded as a line, without direction, joining the points x and y. Those x and y are called the *endpoints* of e; in this case, we say that x and y are *adjacent* to the edge e. The number of edges to which a vertex x of G is adjacent is called the *degree* of x and is denoted by $\deg(x)$. Any edge e with $|\iota(e)| = 1$, that is, any edge e whose two endpoints are equal, is called a *loop*. When two edges e_1, e_2 satisfy that $\iota(e_1) = \iota(e_2)$, we say that e_1 and e_2 are *parallel*, e_1 and e_2 are *multiple edges*, etc. The left part of Figure 5 shows an example of an undirected graph having 5 vertices and 9 edges, where e_1 is a loop ($\iota(e_1) = \{x, x\} = \{x\}$) and e_2 and e_3 are parallel edges ($\iota(e_2) = \iota(e_3) = \{y, z\}$). We say that a graph is *simple* if it has no loops nor multiple edges. For a simple undirected graph, the map ι is injective and consequently, an edge e can be identified with a subset $\iota(e)$ of V. In the following, we often perform such identification without mentioning.

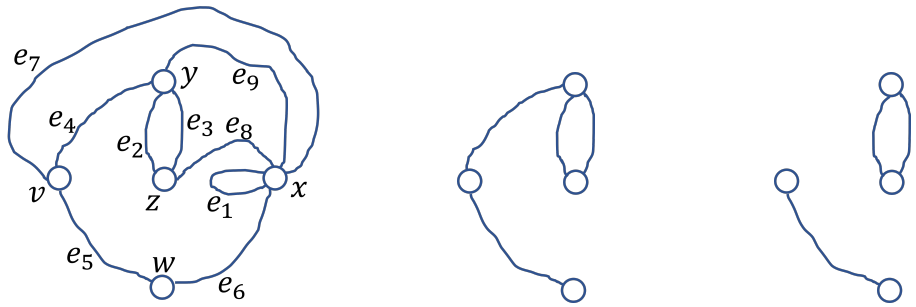


Figure 5: Example of an undirected graph and its subgraphs

Let $G = (V, E, \iota)$ be an undirected graph. We say that a graph is a *subgraph* of G if it is expressed as a form $(V', E', \iota|_{E'})$ where V' is some subset of V and E' is some subset of E'. If moreover this subgraph satisfies that $E' = \{e \in E \mid \iota(e) \subseteq V'\}$, we say that this subgraph is an *induced subgraph*. Induced subgraphs are sometimes called *full subgraphs*. For example, the middle part of Figure 5 is an induced subgraph of the graph in the left part of Figure 5. On the other hand, the right part of Figure 5 is a subgraph of the graph in the left part of Figure 5, but is not an induced subgraph.

Let $G = (V, E, \iota)$ be an undirected graph. A *path* in G is defined as a sequence $v_0e_1v_1e_2 \cdots v_{n-1}e_nv_n$ $(n \geq 0)$ satisfying that for each index i, we have $v_i \in V$, $e_i \in E$, and $\iota(e_i) = \{v_{i-1}, v_i\}$. This path is also expressed by $e_1e_2 \cdots e_n$. Intuitively, this path is formed by visiting vertices $v_0, v_1, \ldots, v_n$ in this order, by passing edges $e_1, e_2, \ldots, e_n$. When such a path satisfies that $v_n = v_0$, we call the path a *closed path* or a *circuit*. For example, in the graph at the left part of Figure 5, $P_1 := ye_4ve_5w$ and $P_2 := ye_2ze_3ye_4ve_4y$ are paths, and $P_3 := ve_5we_6xe_7v$ and $P_4 := ve_7xe_8ze_3ye_9xe_6we_5v$ are closed paths. For a path $v_0e_1v_1e_2 \cdots v_{n-1}e_nv_n$, when the vertices v_i involved are all distinct, we say that this path is *simple*. On the other hand, for a closed path $v_0e_1v_1e_2 \cdots v_{n-1}e_nv_n$, when the vertices v_i involved are all distinct except for the pair $v_n = v_0$, we say that this closed path is *simple*. In the example above, the path P_1 and the closed path P_3 are simple, while the path P_2 and the closed path P_4 are not simple. We note that in a simple graph, an edge

is uniquely determined (if it exists) by the endpoints; in this case, we often write a path $v_0e_1v_1e_2\cdots v_{n-1}e_nv_n$ more simply as $v_0v_1\cdots v_{n-1}v_n$.

For an undirected graph G, we define a relation $x \sim y$ for its vertices x, y to be the existence of a path from x to y. Then this $\sim$ forms an equivalence relation on the vertex set of G. For each of its equivalence class V', the induced subgraph of G with vertex set V' is called a *connected component* of G. When G has at most one connected component, we say that G is *connected*.

For an undirected graph G, we say that G is a *forest* if G has no simple closed path of positive length (the number of edges involved). A *tree* is defined as a connected forest. The graph in the left part of Figure 6 is connected but not a forest. The graph in the middle part of Figure 6 is a forest but not connected (having 2 connected components). The graph in the right part of Figure 6 is a connected forest, i.e., a tree. We note that any forest is a simple graph. For a subgraph G' of G, we say that G' is a *spanning tree* of G if G' is a tree and the vertex set of G' is the whole vertex set of G. For example, the graph in the right part of Figure 6 is a spanning tree of the graph in the left part of Figure 6.

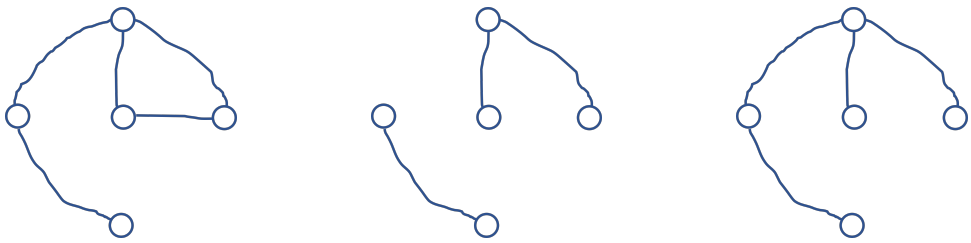


Figure 6: Examples of connected graphs and forests

In contrast to undirected graphs, we also consider situations where each edge of a graph has "direction".

Definition 6.2. A *directed graph* is defined as a triplet (V, E, ι) of two sets V, E and a map $\iota\colon E \to V \times V$.

Similarly to the case of undirected graphs, for a directed graph, we often omit the symbol ι and write (V, E). We say that a directed graph is finite if its vertex set V and its edge set E are both finite.

Intuitively, when vertices $x, y \in V$ and a *directed edge* $e \in E$ (or simply, an edge) of G satisfy that $\iota(e) = (x, y)$, the edge e can be regarded as a line joining points x and y and having direction from x to y. When $\iota(e) = (x, y)$, we call x and y the *source* and the *terminal* of the edge e, denoted by $s(e)$ and $t(e)$, respectively. For a vertex x of G, the number of edges e with $s(e) = x$ is called the *out-degree* of x, and the number of edges e with $t(e) = x$ is called the *in-degree* of x. We say that an edge e is a *loop* if $s(e) = t(e)$. For any edges e_1, e_2 with $\iota(e_1) = \iota(e_2)$, we say that e_1 and e_2 are *parallel*, or *multiple edges*. Any directed graph with no loops nor multiple edges is called *simple*. For any simple directed graph, ι is injective and consequently, an edge e can be identified with the pair $\iota(e)$ of vertices. In the following, we sometimes perform such identification without mentioning. For example, a Hasse diagram of a poset, where each edge is endowed with direction from bottom to top, is a simple directed graph. The notion of subgraphs of a directed graph is defined similarly to the case of undirected graphs.

Let $G = (V, E, \iota)$ be a directed graph. A *directed path* (or simply, a path) in G is defined as a sequence $v_0 e_1 v_1 e_2 \cdots v_{n-1} e_n v_n$ $(n \geq 0)$ satisfying that for each index i, we have $v_i \in V$, $e_i \in E$, and $\iota(e_i) = (v_{i-1}, v_i)$. We also write this path as $e_1 e_2 \cdots e_n$. When such a path satisfies that $v_n = v_0$, we call the path a (directed) *closed path* or a (directed) *circuit*. The notions of simple paths and simple closed paths are defined similarly to the case of undirected graphs. Again, similarly to the case of undirected graphs, in a simple directed graph, we often write such a path as above more simply as $v_0 v_1 \cdots v_{n-1} v_n$. We say that G is *strongly connected* if for any vertices $x, y \in V$ of G, there exists a directed path from x to y.

Definition 6.3. Let $G = (V, E, \iota)$ be a finite simple undirected graph.

- We say that G is a *bipartite graph* if there exists a partition $V = A \sqcup B$

of V satisfying that any edge of G joins some vertex in A and some vertex in B.

- We say that $U \subseteq V$ is a *vertex cover* of G if for any edge e of G we have $\iota(e) \cap U \neq \emptyset$.
- We say that $M \subseteq E$ is a *matching* of G if for any distinct elements e_1, e_2 of M we have $\iota(e_1) \cap \iota(e_2) = \emptyset$. If moreover $\bigcup_{e \in M} \iota(e) = V$ holds, then we say that this matching M is *perfect*.

The following property holds for the sizes of vertex covers and matchings in a bipartite graph. Such a kind of theorems are sometimes categorized as "min-max" theorems.

Theorem 6.1. *For any finite bibartite graph $G = (V, E)$, we have*

$$\min\{|U| \colon U \text{ is a vertex cover of } G\} = \max\{|M| \colon M \text{ is a matching of } G\} \ .$$

Proof. First, for any vertex cover U and any matching M, by the definition of vertex covers, for any $e \in M$ we have $\iota(e) \cap U \neq \emptyset$. By taking an element from this intersection, we obtain a map $f \colon M \to U$. Moreover, the definitions of matchings and the map f imply that f is injective. Hence we have $|M| \leq |U|$. As this inequality holds for any U and M, the right-hand side of the claim is less than or equal to the left-hand side. Hence our task is reduced to showing that the left-hand side of the claim is less than or equal to the right-hand side.

Let M be a matching of G with the largest number of elements. Let $V = A \sqcup B$ be a partition as in the definition of bipartite graphs. Now we define B' to be the set of all vertices $v_{2k+1} \in B$ for which a non-backtracking path (i.e., $e_i \neq e_{i-1}$) as in the following exists:

(*) $v_0 e_1 v_1 \cdots e_{2k+1} v_{2k+1}$, where $v_0 \in A$ is not adjacent to any edge in M, and $e_{2i} \in M$ for any integer i.

We define A' to be the set of all vertices in A that is an endpoint of an edge in M not adjacent to any vertex in B'. In the following, we show that $A' \cup B'$ is a vertex cover of G and satisfies that $|A' \cup B'| = |M|$. Once this is proved, the left-hand side of the claim will be less than or equal to $|A' \cup B'| = |M|$, hence the claim will follow.

For the purpose, we show the following claim.

Claim. *Any vertex in B' is adjacent to some edge in M.*

Proof of Claim. Assume for the contrary that $x \in B'$ is not adjacent to any edge in M. By the definition of B', there is a path $v_0 e_1 v_1 \cdots e_{2k+1} v_{2k+1}$ as in Condition (*) with $v_{2k+1} = x$. Now the assumption above and Condition (*) imply that each of v_0 and v_{2k+1} is not adjacent to any edge in M. Moreover, by the definition of matchings, for any i, each of v_{2i-1} and v_{2i} is not adjacent to any edge in M except for e_{2i}. In particular, it follows that $e_1, e_3, \ldots, e_{2k+1} \notin M$ and that $v_0, \ldots, v_{2k+1}$ are all distinct vertices. By these conditions, the set M' obtained from M by removing $e_2, e_4, \ldots, e_{2k}$ and adding $e_1, e_3, \ldots, e_{2k+1}$ becomes again a matching of G. As $|M'| = |M| + 1$, this contradicts the maximality of M. Hence the current claim holds. □

For each edge e in M, when $\iota(e) \cap B' \neq \emptyset$, we define $g(e)$ to be its unique element (note that as G is a bipartite graph and $B' \subseteq B$, e does not join no two vertices in B'). On the other hand, when $\iota(e) \cap B' = \emptyset$, the endpoint of e that belongs to A (which is uniquely determined, as G is a bipartite graph) also belongs to A' by the definition of A'; we define $g(e)$ to be this vertex. For this map $g\colon M \to A' \cup B'$, as M is a matching, it follows that g is injective; while the definition of A' and the claim above imply that g is surjective. Hence g is bijective, therefore $|A' \cup B'| = |M|$. Now our task is reduced to showing that $A' \cup B'$ is a vertex cover of G.

For the purpose, we assume for the contrary that there is an edge e joining $a \in A \setminus A'$ and $b \in B \setminus B'$. As $b \notin B'$, the path aeb fails Condition (*), therefore a is adjacent to an edge in M, say e'. Let $\iota(e') = \{a, b'\}$ ($b' \in B$). Then the fact $a \notin A'$ implies that $b' \in B'$. Therefore there is a path $v_0 e_1 \cdots e_{2k+1} v_{2k+1}$ with $v_{2k+1} = b'$ satisfying Condition (*). Now the concatenation of the path followed by another path $b'e'aeb$ is also a path satisfying Condition (*), contradicting the fact $b \notin B'$. Hence the claim of this theorem holds. □

We explain a theorem on the condition for existence of perfect matchings in a finite bipartite graph (precisely, such a theorem is deduced by applying the following theorem to the special case $|A| = |B|$).

Theorem 6.2 (Hall's Marriage Theorem)**.** *Let G be a finite bipartite graph and let $V = A \sqcup B$ be the corresponding partition. Then the followings are equivalent:*

1. *There exists a matching M of G for which any vertex in A is adjacent to some edge in M.*
2. *For any subset S of A, when writing the set of vertices joined by an edge to a vertex in S as $N(S)$, we have $|N(S)| \geq |S|$.*

Proof. [1 ⇒ 2] For each $v \in S$, Condition 1 implies that there is an edge e in M for which one of the endpoints is v, and such an e is unique by the definition of matchings. We define $f(v)$ to be the vertex of this e other than v. Then f is a map from S to $N(S)$, and f is injective by the definition of matchings. Hence we have $|S| \leq |N(S)|$, as desired.

[2 ⇒ 1] By the definition of bipartite graphs, a matching M of G with the largest number of elements has at most $|A|$ elements. It suffices to show that $|M| = |A|$; assume for the contrary that $|M| < |A|$. By Theorem 6.1, there is a vertex cover U of G with $|U| < |A|$. Put $A' := U \cap A$ and $B' := U \cap B$. Then as $|U| = |A'| + |B'| < |A|$, we have $|B'| < |A \setminus A'|$. On the other hand, as U is

a vertex cover, we have $N(A \setminus A') \subseteq B'$ and consequently $|N(A \setminus A')| \leq |B'|$. Therefore we have $|N(A \setminus A')| < |A \setminus A'|$, contradicting Condition 2. Hence the claim holds. □

Definition 6.4. We call any pair (T, r) of a tree T and its vertex r a *rooted tree*, and call the r its *root*.

For a given rooted tree, its vertex set is naturally endowed with a partial order. The detail is as follows.

Lemma 6.1. *For any two vertices x, y of a tree T, there exists a unique simple path between x and y.*

Proof. First, as T is connected by definition, there is a path between x and y, and any such path of the shortest length is simple. Now the remaining task is to deduce a contradiction by assuming that there are distinct simple paths $P_1 := z_0 z_1 \cdots z_n$ and $P_2 := w_0 w_1 \cdots w_m$ ($z_0 = w_0 = x$, $z_n = w_m = y$) between x and y. We moreover assume by symmetry that $n \leq m$.

As $z_0 = w_0$, there is the largest index k with $z_k = w_k$. Note that $P_1 \neq P_2$, and as P_2 is simple, for any $i < m$ we have $w_i \neq y$. Therefore we have $k < n \leq m$, while the definition of k implies that $z_{k+1} \neq w_{k+1}$. Put $v := z_k = w_k$ and take a pair of non-negative integers $(\ell_1, \ell_2) \neq (0, 0)$ satisfying $z_{k+\ell_1} = w_{k+\ell_2}$ in a way that $\ell_1 + \ell_2$ is minimum (note that such a pair indeed exists, as $z_n = w_m$). Put $u := z_{k+\ell_1} = w_{k+\ell_2}$. Then both $z_k z_{k+1} \cdots z_{k+\ell_1}$ and $w_k w_{k+1} \cdots w_{k+\ell_2}$ are simple paths from v to u, and by the minimality of $\ell_1 + \ell_2$, both paths do not have a common vertex except for the initial point v and the end point u. Therefore $v z_{k+1} \cdots z_{k+\ell_1-1} u w_{k+\ell_2-1} \cdots w_{k+1} v$ forms a simple closed path in T, contradicting the fact that T is a tree. Hence the claim holds. □

Proposition 6.1. *Let (T, r) be a rooted tree. For two vertices x, y of T, we write the simple path in T from x to y as $P_{x,y}$. Define a relation $x \preceq y$ for vertices x, y as "$P_{r,y}$ involves x". Then $\preceq$ is a partial order on the vertex set $V(T)$ of T.*

Proof. For reflexivity: $P_{r,x}$ involves x by definition, therefore $x \preceq x$.

For antisymmetry: Suppose that $x \preceq y$ and $y \preceq x$. As $x \preceq y$, $P_{r,y}$ involves x, and the part P' of $P_{r,y}$ from r to x becomes $P_{r,x}$. As $y \preceq x$, this $P_{r,x} = P'$ involves y, which holds only when $x = y$. Hence $x = y$.

For transitivity: Suppose that $x \preceq y$ and $y \preceq z$. As $y \preceq z$, $P_{r,z}$ involves y, and the part of $P_{r,z}$ from r to y becomes $P_{r,y}$. As $x \preceq y$, this $P_{r,y}$ involves x, therefore the original $P_{r,z}$ involves x as well. Hence $x \preceq z$. Therefore the claim holds. □

Definition 6.5. Let G be a simple undirected graph. We say that G is a *complete graph* if any two vertices of G is joined by some edge. We write the complete graph with n vertices as K_n.

The number of spanning trees in a complete graph is determined as follows.

Theorem 6.3. *For $n \geq 2$, the number of spanning trees in K_n is n^{n-2}.*

From now on, we give a bijective proof for Theorem 6.3. Let $\mathcal{T}[V]$ be the set of all spanning trees in a complete graph with vertex set $V = \{a_1, \ldots, a_n\} \subseteq \mathbb{Z}$ ($n \geq 2$), and for each $T \in \mathcal{T}[V]$, let $v(T)$ be the first (with respect to the ordering of $\mathbb{Z}$) leaf vertex (i.e., vertex of degree 1) of T. On the other hand, for any $s = (s_1, \ldots, s_{n-2}) \in V^{n-2}$, let $w_V(s)$ be the first element of $V \setminus \{s_1, \ldots, s_{n-2}\}$. Now we define maps $f_V \colon \mathcal{T}[V] \to V^{n-2}$ and $g_V \colon V^{n-2} \to \mathcal{T}[V]$ recursively for $n \geq 2$ as follows (note that when $n = 2$, as the only element of $\mathcal{T}[V]$ is the complete graph on V itself and the only element of V^{n-2} is the empty sequence, the desired correspondence is automatically determined):

- When $T \in \mathcal{T}[V]$, we define $s_1(T)$ to be the unique vertex in V adjacent to $v(T)$ in T, and define $f_V(T)$ to be the composition of $s_1(T)$ followed by the sequence $f_{V \setminus \{v(T)\}}(T \setminus \{v(T)\})$. Here, for any graph G and any subset S of its vertex set $V(G)$, we write $G \setminus S$ to mean the graph

obtained from G by removing S, or more precisely, its induced subgraph with vertex set $V(G) \setminus S$.

- When $s = (s_1, \ldots, s_{n-2}) \in V^{n-2}$, we define $g_V(s)$ to be the tree obtained from the tree $g_{V \setminus \{w_V(s)\}}(s_2, \ldots, s_{n-2})$ by adding a new vertex $w_V(s)$ and joining it to vertex s_1. (Note that we have $w_V(s) \notin \{s_1, \ldots, s_{n-2}\}$ and hence $s_1 \in V \setminus \{w_V(s)\}$.)

Once we show that both $f_V \circ g_V$ and $g_V \circ f_V$ are identity maps, it will follow that f_V is a bijection and consequently, we will have $|\mathcal{T}[V]| = |V^{n-2}| = n^{n-2}$ and hence the claim of Theorem 6.3 will hold. The current claim is obvious when $n = 2$; from now on, we suppose that the current claim holds for smaller n's and prove the current claim when $n \geq 3$.

Lemma 6.2. *In the setting above, if $T \in \mathcal{T}[V]$ and $f_V(T) = (s_1, \ldots, s_{n-2})$, then $\{s_1, \ldots, s_{n-2}\}$ coincides with the set of vertices of degree at least 2 in T, and we have $v(T) = w_V(f_V(T))$.*

Proof. If some s_i is a leaf vertex of T, then by the definition of f_V, by focusing on the tree, say T', at the step of appending s_i to $f_V(T)$, the leaf vertex $v(T')$ of T' was adjacent to s_i. Now as s_i has degree at most 1 in T', the connectedness of trees implies that T' should be the graph consisting of an edge joining s_i and $v(T')$ only. On the other hand, the definition of f_V implies that T' should have at least 3 vertices. This is a contradiction. Hence all the s_i's have degrees at least 2.

Conversely, let v be any vertex of T with degree at least 2, and let u_1, u_2 be distinct vertices adjacent to v. Then during the process of removing the leaf vertices of the tree in the recursive construction of f_V, as the number of vertices becomes finally 2, at least one of u_1 and u_2 must be removed (if both u_1 and u_2 remain not removed, then v does not become a leaf vertex and hence is not removed yet, yielding at least 3 remaining vertices). Moreover, for the tree, say T', at the step of removing u_i ($i \in \{1, 2\}$), we have $v(T') = u_i$, therefore the vertex v that is adjacent to u_i is appended to the sequence

$f_V(T)$. Hence v appears in the sequence $f_V(T)$. This implies the former part of the claim. The latter part of the claim follows from the former part of the claim and the definitions of $v(T)$ and $w_V(s)$. □

Lemma 6.3. *In the setting above, for any $s = (s_1, \ldots, s_{n-2}) \in V^{n-2}$, we have $v(g_V(s)) = w_V(s)$.*

Proof. Let X be the set of all leaf vertices of $g_V(s)$, and let Y be the set of all leaf vertices of $g_{V \setminus \{w_V(s)\}}(s_2, \ldots, s_{n-2})$. By the recursive construction of g_V, we have

$$X = \{w_V(s)\} \sqcup (Y \setminus \{s_1\}) \ .$$

Now the current assumption implies that

$$f_{V \setminus \{w_V(s)\}}(g_{V \setminus \{w_V(s)\}}(s_2, \ldots, s_{n-2})) = \{s_2, \ldots, s_{n-2}\} \ ,$$

therefore Lemma 6.2 implies that

$$Y = (V \setminus \{w_V(s)\}) \setminus \{s_2, \ldots, s_{n-2}\} \ .$$

Consequently, if $u \in Y \setminus \{s_1\}$, then u is an element of $V \setminus \{s_1, \ldots, s_{n-2}\}$ different from $w_V(s)$, therefore the definition of $w_V(s)$ implies that $w_V(s) < u$. Hence we have $v(g_V(s)) = w_V(s)$ and the claim holds. □

Let $T \in \mathcal{T}[V]$ and $f_V(T) = (s_1, \ldots, s_{n-2})$. By definition, $g_V(f_V(T))$ is the graph obtained from $g_{V \setminus \{w\}}(s_2, \ldots, s_{n-2})$, where $w = w_V(f_V(T))$, by adding the first element w of $V \setminus \{s_1, \ldots, s_{n-2}\}$ as a new vertex joined to the vertex s_1. By Lemma 6.2 we have $w = v(T)$, and the recursive construction of f_V and the current assumption imply that

$$g_{V \setminus \{w\}}(s_2, \ldots, s_{n-2}) = g_{V \setminus \{w\}}(f_{V \setminus \{w\}}(T \setminus \{w\})) = T \setminus \{w\} = T \setminus \{v(T)\} \ .$$

This implies that $g_V(f_V(T))$ is the graph obtained from $T \setminus \{v(T)\}$ by adding a new vertex $v(T)$ joined to the vertex s_1, which coincides with T itself by the construction of f_V. Hence we have $g_V(f_V(T)) = T$. Therefore we have $g_V \circ f_V = \mathsf{id}$.

Conversely, let $s = (s_1, \ldots, s_{n-2}) \in V^{n-2}$ and $T := g_V(s)$. Let s'_1 denote the unique vertex of T adjacent to $v(T)$. Then by definition, $f_V(T)$ is the composition of s'_1 followed by the sequence $f_{V\setminus\{v(T)\}}(T \setminus \{v(T)\})$. Lemma 6.3 implies that $v(T) = w_V(s)$, and the construction of g_V implies that $s'_1 = s_1$ and $g_V(s) \setminus \{w_V(s)\} = g_{V\setminus\{w_V(s)\}}(s_2, \ldots, s_{n-2})$. This and the current assumption imply that

$$f_{V\setminus\{v(T)\}}(g_V(s)\setminus\{v(T)\}) = f_{V\setminus\{w_V(s)\}}(g_{V\setminus\{w_V(s)\}}(s_2, \ldots, s_{n-2})) = (s_2, \ldots, s_{n-2}) \ ,$$

therefore

$$f_V(g_V(s)) = f_V(T) = (s'_1, s_2, \ldots, s_{n-2}) = (s_1, s_2, \ldots, s_{n-2}) = s \ .$$

Hence we have $f_V \circ g_V = \mathsf{id}$. This completes the proof of Theorem 6.3.

Exercises

Problem 1. Prove that if a finite undirected graph $G = (V, E)$ has no loops, then we have $\sum_{v\in V} \deg(v) = 2|E|$.

(Hint: Use a counting argument.)

Problem 2. Prove that any finite tree with at least 2 vertices has a vertex of degree 1.

(Comment: A counterexample exists for the case of infinite trees. For example, consider the graph with the integers as vertices and the edges joining consecutive integers.)

Problem 3. For any finite tree $G = (V, E)$ with non-empty vertex set, prove that $|E| = |V| - 1$.

(Hint: Use the previous problem and mathematical induction.)

Problem 4. Prove that any connected undirected graph with non-empty vertex set has a spanning tree.

(Hint: For the case of infinite vertex set, use Zorn's Lemma.)

7 Ramsey Theory

Definition 7.1. For a simple undirected graph $G = (V, E)$, its *complement graph* $\overline{G} = (V', E')$ is defined by $V' := V$ and

$$E' := \{\{x, y\} \mid x, y \in V, x \neq y, \{x, y\} \notin E\} \ .$$

For example, the complement graph $\overline{K_r}$ of the complete graph K_r with r vertices is the graph with r vertices and no edges.

Definition 7.2. For any finite undirected graph, we say that its induced subgraph is a *clique* if it is of the form K_r for some r, and an *independent set* if it is of the form $\overline{K_r}$ for some r.

In graph theory, a kind of theorems like "any sufficiently large graph has a certain special substructure" have been well studied. We explain some examples of them.

Theorem 7.1. *Let $r \geq 2$ be an integer. Then any simple undirected graph $G = (V, E)$ with at least 2^{2r-3} vertices has either a clique with r vertices or an independent set with r vertices.*

Proof. It suffices to prove the claim for the case $|V| = 2^{2r-3}$. We set $V_1 := V$, $I, J := \emptyset$, and take some $v_1 \in V_1$. For $i = 2, \ldots, 2r-2$, we choose $V_i \subseteq V_{i-1}$ with $|V_i| = 2^{2r-2-i}$ and $v_i \in V_i$ recursively, while updating the sets I and J, in the following manner (we note that for the case $i = 1$ we also have $|V_1| = 2^{2r-2-1}$ and $v_1 \in V_1$):

- We note that there are $2^{2r-2-(i-1)} - 1$ vertices in V_{i-1} other than v_{i-1}, therefore one of the followings holds.
  - At least $2^{2r-3-(i-1)} = 2^{2r-2-i}$ vertices in V_{i-1} are adjacent to v_{i-1}.
  - At least $2^{2r-3-(i-1)} = 2^{2r-2-i}$ vertices in V_{i-1} are not adjacent to v_{i-1}.

In the former case, we let V_i consist of any 2^{2r-2-i} vertices in V_{i-1} adjacent to v_{i-1}, append $i-1$ to I, and take some $v_i \in V_i$. In the latter case, we let V_i consist of any 2^{2r-2-i} vertices in V_{i-1} not adjacent to v_{i-1}, append $i-1$ to J, and take some $v_i \in V_i$.

By the construction, we have $I \cup J = \{1, \ldots, 2r-3\}$ and $I \cap J = \emptyset$, therefore either $|I| \geq r-1$ or $|J| \geq r-1$ holds. When $|I| \geq r-1$, any r vertices chosen from $\{v_i \mid i \in I\} \cup \{v_{2r-2}\}$ form a clique of size r. On the other hand, when $|J| \geq r-1$, any r vertices chosen from $\{v_i \mid i \in J\} \cup \{v_{2r-2}\}$ form an independent set of size r. Hence the claim holds. □

Definition 7.3. Let $r \geq 2$ be an integer. We define $R(r)$ to be the minimum n satisfying "any simple undirected graph with n vertices has either a clique of size r or an independent set of size r", and call it the *Ramsey number*.

By Theorem 7.1, the Ramsey number $R(r)$ is well-defined as a finite value, and we have $R(r) \leq 2^{2r-3}$. For example, for $r = 3$ we have $R(3) \leq 2^3 = 8$ (see also the exercise below).

We give a lower bound of the Ramsey number by using a probabilistic argument as follows. We define a random variable $\mathcal{G}(n,p)$ over simple undirected graphs with a given set of n vertices in a way that for any two distinct vertices x, y, they are joined by an edge with probability p independently.

Lemma 7.1. *For any $n \geq k \geq 2$, the $\mathcal{G}(n,p)$ above satisfies*

$$p_1 := \Pr[\mathcal{G}(n,p) \text{ has an independent set of size } k] \leq \binom{n}{k}(1-p)^{\binom{k}{2}} ,$$

$$p_2 := \Pr[\mathcal{G}(n,p) \text{ has a clique of size } k] \leq \binom{n}{k}p^{\binom{k}{2}} .$$

Proof. Let V be the vertex set of $\mathcal{G}(n,p)$. For any $S \subseteq V$, no two vertices in S are joined by an edge with probability $(1-p)^{\binom{|S|}{2}}$, and every two vertices are joined by an edge with probability $p^{\binom{|S|}{2}}$. This implies that

$$p_1 \leq \sum_{S \subseteq V\,;\,|S|=k} \Pr[S \text{ forms an independent set}] = \sum_{S \subseteq V\,;\,|S|=k} (1-p)^{\binom{k}{2}} = \binom{n}{k}(1-p)^{\binom{k}{2}} ,$$

$$p_2 \leq \sum_{S \subseteq V\,;\,|S|=k} \Pr[S \text{ forms a clique}] = \sum_{S \subseteq V\,;\,|S|=k} p^{\binom{k}{2}} = \binom{n}{k} p^{\binom{k}{2}} \ .$$

Hence the claim holds. □

Theorem 7.2. *For any $r \geq 3$, we have $R(r) > 2^{r/2}$.*

Proof. When $r = 3$, it is known that $R(3) = 6$ (see the exercise below), therefore $R(3) > 2^{3/2}$. From now on, we let $r \geq 4$. When $n \leq 2^{r/2}$, Lemma 7.1 implies that

$$\begin{aligned} &\Pr[\mathcal{G}(n, 1/2) \text{ has a clique of size } r] \\ &\leq \binom{n}{r} \left(\frac{1}{2}\right)^{\binom{r}{2}} \\ &= \frac{n(n-1)\cdots(n-r+1)}{r!} 2^{-r(r-1)/2} \\ &< \frac{n^r}{2^r} 2^{-r(r-1)/2} \end{aligned}$$

(here we used the fact $r! > 2^r$ implied by $r \geq 4$)

$$\leq \frac{(2^{r/2})^r}{2^{r(r-1)/2+r}} = \frac{2^{r^2/2}}{2^{r^2/2+r/2}} = \frac{1}{2^{r/2}} < \frac{1}{2} \ ,$$

and we similarly have $\Pr[\mathcal{G}(n, 1/2) \text{ has an independent set of size } r] < 1/2$. This implies that the probability that $\mathcal{G}(n, 1/2)$ has a clique of size r or an independent set of size r is less than $1/2 + 1/2 = 1$, therefore the probability that $\mathcal{G}(n, 1/2)$ has no cliques of size r nor independent sets of size r is positive. Hence, when $n \leq 2^{r/2}$, there is a simple undirected graph with n vertices having no cliques of size r nor independent sets of size r. Therefore we have $R(r) > 2^{r/2}$ and the claim holds. □

We explain another proposition of the form "any sufficiently large graph has a certain special substructure". We define a *complete bipartite graph* $K_{n,m}$ to be a bipartite graph with partition of vertex set $V = A \sqcup B$ satisfying that $|A| = n$, $|B| = m$, and any vertex in A and any vertex in B are joined by an edge.

Proposition 7.1. *Let $r \geq 2$ be an integer. Then there exists an integer N satisfying the following: Any simple connected undirected graph $G = (V, E)$ with at least N vertices has an induced subgraph of one of the following forms; K_r, $K_{1,r}$, or a simple path P_r of length r (i.e., having r edges).*

Proof. We first consider the case where some vertex v of G has degree at least $R(r)$. As the induced subgraph G' of G consisting of the vertices adjacent to v has at least $R(r)$ vertices, the definition of $R(r)$ implies that G' has either K_r or $\overline{K_r}$ as its induced subgraph. In the former case, the K_r satisfies the claim; while in the latter case, the induced subgraph obtained by gathering the $\overline{K_r}$ and v forms the graph $K_{1,r}$ as in the claim. Hence the claim holds in this case.

From now on, we consider the remaining case where every vertex of G has degree at most $R(r) - 1$. Starting from some vertex v and a set $V_0 := \{v\}$, we define sets V_i $(1 \leq i \leq r-1)$ recursively by

$$V_i := \left\{ u \in V \setminus \bigcup_{j<i} V_j \mid u \text{ is adjacent to some vertex in } V_{i-1} \right\} .$$

Now the aforementioned condition for the degrees implies that $|V_i| \leq (R(r)-1)|V_{i-1}|$, therefore we have $|V_i| \leq (R(r)-1)^i$. Hence we have

$$\left| \bigcup_{i=0}^{r-1} V_i \right| \leq \sum_{i=0}^{r-1} (R(r)-1)^i .$$

By writing the right-hand side (independent of G) as M, whenever $|V| \geq M + 1$, there is a $u \in V$ not belonging to any of V_i. As G is connected, there is a simple path P from u and v, and the choice of u implies that P has length at least r. Now the first $r + 1$ vertices of P forms an induced subgraph P_r as in the claim. Hence the claim holds. □

Exercises

Problem 1. For the Ramsey number, prove that $R(3) = 6$.

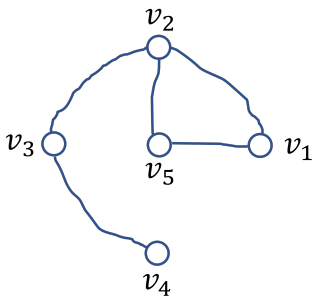
8 Adjacency Matrices of Graphs

In this section, we let $G = (V, E)$ be a finite simple undirected graph.

Definition 8.1. Suppose that the elements of V are enumerated as $\{v_1, \ldots, v_n\}$. We define the *adjacency matrix* of the graph G, denoted by $A = A(G) = (a_{ij})_{i,j=1}^n$, by

$$a_{ij} := \begin{cases} 1 & (\text{if } \{v_i, v_j\} \in E) \ , \\ 0 & (\text{otherwise}) \ . \end{cases}$$

An example of the adjacency matrix of a graph is shown in Figure 7. We note that by definition, the adjacency matrix A is a symmetric matrix and the sum of the components in the i-th row of A is equal to the degree $\deg(v_i)$ of v_i.



$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Figure 7: Example of the adjacency matrix of a graph

As the original graph can be fully recovered from its adjacency matrix, any information on a graph could in principle be deduced from its adjacency matrix (but, of course, such information is not always deduced efficiently). For example, we have the following property.

Proposition 8.1. *Let A be the adjacency matrix of G. Then the (i, j)-entry of A^k coincides with the number of paths of length k in G from v_i to v_j.*

Proof. By the definition of multiplication of matrices, the (i, j)-entry of A^k

can be expressed as

$$\sum_{\substack{\ell_0,\ell_1,\dots,\ell_k \\ \ell_0=i\,,\,\ell_k=j}} a_{\ell_0,\ell_1} a_{\ell_1,\ell_2} \cdots a_{\ell_{k-1},\ell_k} \ .$$

The term $a_{\ell_0,\ell_1} a_{\ell_1,\ell_2} \cdots a_{\ell_{k-1},\ell_k}$ in this sum is equal to 1 if all of $\{v_{\ell_0}, v_{\ell_1}\}$, $\{v_{\ell_1}, v_{\ell_2}\}, \dots, \{v_{\ell_{k-1}}, v_{\ell_k}\}$ are edges, or equivalently $v_{\ell_0} v_{\ell_1} \cdots v_{\ell_{k-1}} v_{\ell_k}$ is a path; and it is 0 otherwise. Therefore, by concerning the conditions $v_{\ell_0} = v_i$ and $v_{\ell_k} = v_j$ in the sum, it follows that the sum above is equal to the number of paths of length k from v_i to v_j. Hence the claim holds. □

Corollary 8.1. *Let A be the adjacency matrix of G. Then the (i,i)-entry of A^2 is equal to* $\deg(v_i)$.

Proof. As G is a simple graph, any path of length 2 from v_i to v_i is of the form $v_i u v_i$ with u being a vertex adjacent to v_i. This and Proposition 8.1 imply the claim. □

Definition 8.2. In the setting of Definition 8.1, let D denote the $n \times n$ matrix with the diagonal entries being $\deg(v_1), \dots, \deg(v_n)$ and the other entries being 0. Then we call $D - A$ the *Laplacian matrix* of G.

Proposition 8.2. *We enumerate the edges of G as $E = \{e_1, \dots, e_m\}$ and associate some direction to each edge; that is, for each edge, we let some of the two endpoints be "the source" and the other endpoint be "the terminal". Moreover, we define an $n \times m$ matrix $B = (B_{ij})_{i,j}$ by*

$$b_{ij} = \begin{cases} 1 & \textit{(if } v_i \textit{ is the source of } e_j\textit{)} \\ -1 & \textit{(if } v_i \textit{ is the terminal of } e_j\textit{)} \\ 0 & \textit{(otherwise)} \ . \end{cases}$$

Then we have $BB^{\mathrm{T}} = D - A$. In particular, $D - A$ is a positive-semidefinite symmetric matrix.

Proof. First, the (i,i)-entry of BB^{T} can be expressed as

$$\sum_j b_{ij} b_{ij} = \sum_j {b_{ij}}^2 \ .$$

For each term in the sum, if v_i is an endpoint of e_j then we have ${b_{ij}}^2 = 1$ (regardless of the direction of the edge), and otherwise we have ${b_{ij}}^2 = 0$. Therefore, the sum involves the same number of terms 1 as the number of the edges adjacent to v_i, resulting in the sum being $\deg(v_i)$. As G is a simple graph, the diagonal entries of A are all 0, therefore the sum above is equal to the (i,i)-entry of $D - A$.

Secondly, when $i \neq j$, the (i,j)-entry of BB^{T} can be expressed as

$$\sum_k b_{ik} b_{jk} \ .$$

For each term in the sum, if e_k is an edge from v_i to v_j then we have $b_{ik}b_{jk} = 1 \cdot (-1) = -1$; if e_k is an edge from v_j to v_i then we have $b_{ik}b_{jk} = (-1) \cdot 1 = -1$; and otherwise we have $b_{ik}b_{jk} = 0$. Now as G is a simple graph, the sum above becomes -1 if v_i and v_j are adjacent to each other, and otherwise it becomes 0. This value is equal to the (i,j)-entry of $D - A$. Hence the claim holds. □

The following holds for the number of spanning trees in a graph.

Theorem 8.1 (Kirchhoff's Matrix-Tree Theorem)**.** *The (i,i)-cofactor of the Laplacian matrix $D - A$ of G is equal to the number of spanning trees in G.*

Proof. First we note that, when re-ordering the vertices by moving v_i to the last, the (n,n)-cofactor of the resulting Laplacian matrix coincides with the (i,i)-cofactor of the original Laplacian matrix. Therefore, we may assume without loss of generality that $i = n$.

We write $D - A = BB^{\mathrm{T}}$ as in Proposition 8.2. Let $B' = (b_{ij})_{1 \leq i \leq n-1 \, , \, 1 \leq j \leq m}$ be the matrix obtained from B by removing the last row. Then the (n,n)-cofactor of $D - A$ can be expressed as $\det(B'B'^{\mathrm{T}})$. As the (i,j)-entry of

$B'B'^{\mathrm{T}}$ is $\displaystyle\sum_{k=1}^{m} b_{ik}b_{jk}$, the (n,n)-cofactor of $D-A$ is given by

$$\begin{aligned}&\sum_{\sigma\in S_{n-1}} \mathrm{sgn}(\sigma) \sum_{k_1,\dots,k_{n-1}\in\{1,\dots,m\}} b_{1,k_1}b_{\sigma(1),k_1}b_{2,k_2}b_{\sigma(2),k_2}\cdots b_{n-1,k_{n-1}}b_{\sigma(n-1),k_{n-1}}\\ =&\sum_{k_1,\dots,k_{n-1}\in\{1,\dots,m\}} \sum_{\sigma\in S_{n-1}} \mathrm{sgn}(\sigma) b_{1,k_1}b_{\sigma(1),k_1}b_{2,k_2}b_{\sigma(2),k_2}\cdots b_{n-1,k_{n-1}}b_{\sigma(n-1),k_{n-1}} \ . \end{aligned} \tag{3}$$

We write the term $b_{1,k_1}b_{\sigma(1),k_1}b_{2,k_2}b_{\sigma(2),k_2}\cdots b_{n-1,k_{n-1}}b_{\sigma(n-1),k_{n-1}}$ as P_σ. Note that P_σ becomes ± 1 when for every $i\in\{1,\dots,n-1\}$, each of v_i and $v_{\sigma(i)}$ is an endpoint of e_{k_i}; otherwise P_σ becomes 0. In the following, we focus on the case where $P_\sigma\neq 0$.

We suppose that $i\neq j$ and $k_i=k_j$. If e_{k_i} is not an edge joining v_i and v_j, then we have $P_\sigma=0$. In the following, we consider the other case where e_{k_i} is an edge joining v_i and v_j. In this case, $P_\sigma\neq 0$ holds only when $\sigma(i)\in\{i,j\}$ and $\sigma(j)\in\{i,j\}$, or equivalently either $(\sigma(i),\sigma(j))=(i,j)$ or $(\sigma(i),\sigma(j))=(j,i)$. Now the σ's satisfying the former condition and the σ's satisfying the latter condition are in one-to-one correspondence via the rule $\sigma\mapsto(i,j)\sigma$. Moreover, as $b_{i,k_i}b_{i,k_i}b_{j,k_i}b_{j,k_i}=b_{i,k_i}b_{j,k_i}b_{j,k_i}b_{i,k_i}$, we have $P_{(i,j)\sigma}=P_\sigma$. This and the fact $\mathrm{sgn}((i,j)\sigma)=-\mathrm{sgn}(\sigma)$ imply that in the sum $\displaystyle\sum_{\sigma\in S_{n-1}}\mathrm{sgn}(\sigma)P_\sigma$, the term corresponding to σ and the term corresponding to $(i,j)\sigma$ are cancelled with each other; therefore the sum above becomes 0. Consequently, in the right-hand side of Eq.(3), the inner sum becomes 0 unless the k_i's are all distinct. In the following, we focus on the remaining case where the k_i's are all distinct.

We suppose that $i_1,i_2,\dots,i_\ell$ are all distinct and the edges $e_{k_{i_1}},e_{k_{i_2}},\dots,e_{k_{i_\ell}}$ in this order form a simple closed path (in particular $\ell\geq 3$). Here the ordering of the edges in the simple closed path is chosen in a way that the endpoint v_{i_1} of $e_{k_{i_1}}$ is also an endpoint of $e_{k_{i_\ell}}$ (rather than $e_{k_{i_2}}$). Now it can be recursively shown that for $j=\ell-1,\dots,2,1$, v_{i_j} is a common endpoint of $e_{k_{i_j}}$ and $e_{k_{i_{j+1}}}$. Therefore the simple closed path above is of the form

$v_{i_1}e_{k_{i_1}}v_{i_2}e_{k_{i_2}}\cdots v_{i_\ell}e_{k_{i_\ell}}v_{i_1}$. Put

$$P'_\sigma := b_{i_1,k_{i_1}}b_{\sigma(i_1),k_{i_1}}b_{i_2,k_{i_2}}b_{\sigma(i_2),k_{i_2}}\cdots b_{i_\ell,k_{i_\ell}}b_{\sigma(i_\ell),k_{i_\ell}}\ ,\ P''_\sigma := P_\sigma/P'_\sigma\ .$$

We note that for each j, $v_{\sigma(i_j)}$ is also an endpoint of $e_{k_{i_j}}$. In particular, we have $\sigma(i_1) \in \{i_1, i_2\}$. Now:

- When $\sigma(i_1) = i_1$, we show recursively for $j = \ell, \ell-1, \ldots, 2$ that $\sigma(i_j) = i_j$. For the case $j = \ell$, the fact that $v_{\sigma(i_\ell)}$ is an endpoint of $e_{k_{i_\ell}}$ implies that $\sigma(i_\ell) \in \{i_\ell, i_1\}$, therefore the fact $\sigma(i_\ell) \neq \sigma(i_1) = i_1$ implies that $\sigma(i_\ell) = i_\ell$, as desired. Moreover, assuming that this claim holds for $j+1$ and the previous steps, the fact that $v_{\sigma(i_j)}$ is an endpoint of $e_{k_{i_j}}$ implies that $\sigma(i_j) \in \{i_j, i_{j+1}\}$, therefore the fact $\sigma(i_j) \neq \sigma(i_{j+1}) = i_{j+1}$ implies that $\sigma(i_j) = i_j$, as desired.

- When $\sigma(i_1) = i_2$, we show recursively for $j = 1, 2, \ldots, \ell$ that $\sigma(i_j) = i_{j+1}$ (where we put $i_{\ell+1} := i_1$). This claim holds immediately for $j = 1$. Moreover, assuming that this claim holds for $j-1$ and the previous steps, the fact that $v_{\sigma(i_j)}$ is an endpoint of $e_{k_{i_j}}$ implies that $\sigma(i_j) \in \{i_j, i_{j+1}\}$, therefore the fact $\sigma(i_j) \neq \sigma(i_{j-1}) = i_j$ implies that $\sigma(i_j) = i_{j+1}$, as desired.

By putting $\tau := (i_1, i_2, \ldots, i_\ell) \in S_n$, the former σ's and the latter σ's are in one-to-one correspondence via the rule $\sigma \mapsto \tau\sigma$, and we have $P''_\sigma = P''_{\tau\sigma}$. Now for the former σ, we have $P'_\sigma = {b_{i_1,k_{i_1}}}^2 \cdot {b_{i_2,k_{i_2}}}^2 \cdots {b_{i_\ell,k_{i_\ell}}}^2 = 1$; while for the latter $\tau\sigma$, we have

$$\begin{aligned} P'_{\tau\sigma} &= (b_{i_1,k_{i_1}}b_{i_2,k_{i_1}}) \cdot (b_{i_2,k_{i_2}}b_{i_3,k_{i_2}}) \cdots (b_{i_\ell,k_{i_\ell}}b_{i_1,k_{i_\ell}}) \\ &= (-1) \cdot (-1) \cdots (-1) = (-1)^\ell \end{aligned}$$

and $\mathrm{sgn}(\tau\sigma) = \mathrm{sgn}(\tau)\mathrm{sgn}(\sigma) = (-1)^{\ell-1}\mathrm{sgn}(\sigma)$. Consequently, in the sum $\displaystyle\sum_{\sigma\in S_{n-1}} \mathrm{sgn}(\sigma)P_\sigma$, the term corresponding to σ and the term corresponding to $\tau\sigma$ are cancelled with each other, therefore the sum becomes 0.

Let $\mathcal{G}$ be the set of the sequences $(k_1, \ldots, k_{n-1})$ with distinct components satisfying that the subgraph $G' := (V, \{e_{k_1}, \ldots, e_{k_{n-1}}\})$ does not have a simple closed path. Then the argument above implies that the right-hand side of Eq.(3) becomes

$$\sum_{(k_1,\ldots,k_{n-1})\in\mathcal{G}} \sum_{\sigma\in S_{n-1}} \mathrm{sgn}(\sigma) P_\sigma \ . \tag{4}$$

As $|V| = n$, the condition that G' has no simple closed paths is equivalent to that G' is a spanning tree of G. Now when G' is a spanning tree of G, for each $v \in V \setminus \{v_n\}$, let $\varphi(v)$ denote the last edge of the unique simple path in G' from v_n to v. If $(k_1, \ldots, k_{n-1}) \in \mathcal{G}$ and $P_\sigma \neq 0$, we show that for any $1 \leq i \leq n-1$ we have $e_{k_i} = \varphi(v_i)$ and $\sigma(i) = i$, by mathematical induction on the length of the simple path in G' from v_n to v_i.

- When v_i is adjacent to v_n in G', suppose that $e_{k_j} = \varphi(v_i)$ is the edge joining v_i and v_n. Then the property $P_\sigma \neq 0$ implies that v_j has to be an endpoint of e_{k_j}; as $v_j \neq v_n$, we have $v_j = v_i$, therefore we have $j = i$ and $e_{k_i} = \varphi(v_i)$. Moreover, the property $P_\sigma \neq 0$ also implies that $v_{\sigma(i)}$ has to be an endpoint of e_{k_i} as well; as $v_{\sigma(i)} \neq v_n$, we have $v_{\sigma(i)} = v_i$ and $\sigma(i) = i$.

- In the other case, suppose that v_j is the vertex, right before v_i, of the simple path in G' from v_n to v_i. Then the induction hypothesis implies that $e_{k_j} = \varphi(v_j)$ and $\sigma(j) = j$. By writing $\varphi(v_i) = e_{k_\ell}$, it follows that e_{k_ℓ} is the edge joining v_i and v_j. Now the property $P_\sigma \neq 0$ implies that v_ℓ is an endpoint of e_{k_ℓ}, which is either v_i or v_j; while, as $v_i \neq v_j$, we have $e_{k_\ell} = \varphi(v_i) \neq \varphi(v_j) = e_{k_j}$, therefore $\ell \neq j$. This implies that $v_\ell = v_i$ and $\ell = i$, therefore we have $\varphi(v_i) = e_{k_i}$. Moreover, the property $P_\sigma \neq 0$ also implies that $v_{\sigma(i)}$ is an endpoint of e_{k_i} as well, which is either v_i or v_j; while as $i \neq j$, we have $\sigma(i) \neq \sigma(j) = j$. This implies that $v_{\sigma(i)} = v_i$ and $\sigma(i) = i$.

Consequently, for any fixed spanning tree G', the pair of $(k_1, \ldots, k_{n-1})$ and σ satisfying $P_\sigma \neq 0$ is uniquely determined by the conditions that for any i we

have $\varphi(v_i) = e_{k_i}$ and $\sigma(i) = i$. Moreover, as now we have $\sigma = \mathsf{id}$, it follows that $\mathrm{sgn}(\sigma) P_\sigma = {b_{1,k_1}}^2 \cdots {b_{n-1,k_{n-1}}}^2 = 1$. Summarizing, the value of Eq.(4) is equal to the sum of the same number of 1's as the number of spanning trees G' in G, which is also equal to the number of spanning trees in G. Hence the claim holds. □

Example 8.1. We already explained as Theorem 6.3 that for any $n \geq 2$, the number of spanning trees in $G = K_n$ is n^{n-2}. Here we verify this fact again, by using Theorem 8.1. The Laplacian matrix of K_n is given by

$$\begin{pmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & n-1 \end{pmatrix} \quad \text{(a square matrix of size } n\text{)} \ ,$$

and its $(1,1)$-cofactor is given by

$$\det \begin{pmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & n-1 \end{pmatrix} \quad \text{(a square matrix of size } n-1\text{)} \ .$$

By applying elementary row transformations "subtract the second row from the third to $(n-1)$-th rows" and "add $n-1$ multiple of the second row to the first row" to the matrix above, its determinant becomes

$$\det \begin{pmatrix} 0 & n^2-2n & -n & -n & \cdots & -n \\ -1 & n-1 & -1 & -1 & \cdots & -1 \\ 0 & -n & n & 0 & \cdots & 0 \\ 0 & -n & 0 & n & \cdots & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & -n & 0 & 0 & \cdots & n \end{pmatrix} = \det \begin{pmatrix} n^2-2n & -n & -n & \cdots & -n \\ -n & n & 0 & \cdots & 0 \\ -n & 0 & n & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ -n & 0 & 0 & \cdots & n \end{pmatrix}$$

(a square matrix of size $n-2$). Moreover, by adding the second to $(n-2)$-th rows to the first row, the determinant above becomes (by using $n^2 - 2n -$

$(n-3)n = n$)

$$\det \begin{pmatrix} n & 0 & 0 & \cdots & 0 \\ -n & n & 0 & \cdots & 0 \\ -n & 0 & n & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ -n & 0 & 0 & \cdots & n \end{pmatrix} = n^{n-2}$$

(where we used the property that the last matrix is a lower triangular matrix). Hence the claim holds.

We say that a graph is a *d-regular graph* if every vertex has the constant degree d.

Definition 8.3. For any subset $S \subseteq V$, let ∂S denote the set of all edges joining some vertex in S and some vertex in $V \setminus S$. We call

$$h(G) := \min \left\{ \frac{|\partial S|}{|S|} \colon S \subseteq V \,,\, 0 < |S| \leq \frac{1}{2}|V| \right\}$$

the *expansion constant* of G. Moreover, we say that a family $(G_i = (V_i, E_i))_{i \geq 1}$ of d-regular graphs (with common d) is an *expander family* if $\lim_{n \to \infty} |V_n| = \infty$ and there is a positive lower bound of $\{h(G_n) \mid n = 1, 2, \dots\}$.

We explain some relations between the expansion constant and the adjacency matrix of a graph.

Definition 8.4. For the adjacency matrix A of G, we define $\lambda_k = \lambda_k(G)$ to be the k-th largest eigenvalues of A (counting multiplicity) and call it the *k-th eigenvalue* of G. We also call $\lambda_1(G) - \lambda_2(G)$ the *spectral gap* of G.

We note that the k-th eigenvalue λ_k of a graph G is independent of the ordering over the vertices of G. Indeed, any permutation of vertices changes the adjacency matrix A to $P^{-1}AP$ (where P is some permutation matrix), which preserves the eigenvalues.

Lemma 8.1. *Let G be a d-regular graph. Then we have $\lambda_1(G) = d$.*

Proof. First, as G is d-regular, it follows that $A\vec{1} = d\vec{1}$ where $\vec{1}$ denotes the vector with all components being 1, therefore d is an eigenvalue of G. On the other hand, the fact that G is d-regular also implies that $D = dI$ (where I is the identity matrix), while Proposition 8.2 implies that $D - A$ is a positive-semidefinite symmetric matrix. Therefore all the eigenvalues of $dI - A$ are non-negative real numbers, while $d - \lambda_1(G)$ is an eigenvalue of $dI - A$; therefore we have $d - \lambda_1(G) \geq 0$ and hence $d \leq \lambda_1(G) \leq d$. Therefore the claim holds. □

The following fact is known about relations between the expander constant and the spectral gap of a graph; we omit the proof here.

Theorem 8.2 (Allon–Milman)**.** *If G is a d-regular graph (hence $\lambda_1(G) = d$), then we have*

$$\frac{d - \lambda_2(G)}{2} \leq h(G) \leq \sqrt{2d(d - \lambda_2(G))} \ .$$

For any connected graph G, the *diameter* $\mathrm{diam}(G)$ of G is defined to be the maximum length of the shortest path between two vertices of G. The following fact is known; we omit the proof here.

Theorem 8.3 (Nilli)**.** *If G is a connected d-regular graph with $\mathrm{diam}(G) \geq 2m + 2 \geq 4$, then we have*

$$\lambda_2(G) > 2\sqrt{d-1} - \frac{2\sqrt{d-1} - 1}{m} \ .$$

For any d-regular graph G, we define

$$\lambda(G) := \max\{|\lambda_j(G)| \colon \lambda_j(G) \neq \pm d\} \ .$$

It is known (though we omit the proof here) that for any connected d-regular graph G, if G is a bipartite graph with at least 3 vertices, then we have $\lambda(G) = \lambda_2(G)$; while if G is not a bipartite graph, then for any $2 \leq i \leq n$ (where n is the number of vertices of G) we have $|\lambda_i(G)| < d$, therefore $\lambda(G) = \max\{|\lambda_2(G)|, |\lambda_n(G)|\}$. In particular, we have $\lambda(G) \geq \lambda_2(G)$ in any

case. Based on Theorem 8.3, G is called a Ramanujan graph when $\lambda(G)$ takes "almost the smallest possible" value. More precisely, we have the following definition.

Definition 8.5. Let G be a connected d-regular graph. We say that G is a *Ramanujan graph* if $\lambda(G) \leq 2\sqrt{d-1}$.

At the end of this section, we explain about a relation between random walks on a graph and the spectral gap of the graph. Let G be a connected d-regular graph with n vertices that is not a bipartite graph. In a random walk on G, we start from some vertex of G, and repeat the steps where we move to one of the vertices of G adjacent to the current vertex, with equal probabilities (probability $1/d$). We discuss about the limit distribution and the speed of convergence of the random walk. We enumerate the vertices of G as $v_1, \ldots, v_n$ and start from the vertex v_1. First we note that during the first k steps, any path of length k starting from v_1 arises with equal probabilities. By Proposition 8.1, the number of paths of length k from v_1 to v_j coincides with the $(1,j)$-entry of A^k, which is expressed as ${e_1}^{\mathrm{T}} A^k e_j$ where e_i denotes the i-th coordinate vector. Consequently, by writing the n-dimensional vector with all components being 1 as $\vec{1}$, the total number of paths of length k starting from v_1 becomes ${e_1}^{\mathrm{T}} A^k \vec{1}$. As G is d-regular, this value is equal to ${e_1}^{\mathrm{T}} \cdot d^k \vec{1} = d^k$. Therefore, the probability that we arrive at v_j in the k-th step, denoted by $p_j = p_j^{\langle k \rangle}$, is

$$p_j = \frac{{e_1}^{\mathrm{T}} A^k e_j}{d^k} \ .$$

Now as A is a real symmetric matrix, we can diagonalize it by an orthogonal matrix. That is, there is an orthonormal basis $(u_1, \ldots, u_n)$ of $\mathbb{R}^n$ satisfying that for any i we have $Au_i = \lambda_i(G) u_i$. Then we can write $A = \sum_{i=1}^{n} \lambda_i(G) u_i \cdot {u_i}^{\mathrm{T}}$, therefore we have $A^k = \sum_{i=1}^{n} \lambda_i(G)^k u_i \cdot {u_i}^{\mathrm{T}}$ and hence

$$ {e_1}^{\mathrm{T}} A^k e_j = \sum_{i=1}^{n} \lambda_i(G)^k \cdot {e_1}^{\mathrm{T}} u_i \cdot {u_i}^{\mathrm{T}} e_j \ .$$

As G is d-regular, we have $\lambda_1(G) = d$ and we can set $u_1 = n^{-1/2}\vec{1}$. Therefore we have

$$e_1{}^{\mathrm{T}} A^k e_j = \frac{d^k}{n} + \sum_{i=2}^{n} \lambda_i(G)^k \cdot e_1{}^{\mathrm{T}} u_i \cdot u_i{}^{\mathrm{T}} e_j \ .$$

Hence we have

$$\left| p_j - \frac{1}{n} \right| \leq \sum_{i=2}^{n} \frac{|\lambda_i(G)^k|}{d^k} \cdot |e_1{}^{\mathrm{T}} u_i| \cdot |u_i{}^{\mathrm{T}} e_j| \ .$$

By Cauchy–Schwarz Inequality, the right-hand side becomes

$$\leq \sum_{i=2}^{n} \frac{|\lambda_i(G)|^k}{d^k} \cdot ||e_1|| \cdot ||u_i|| \cdot ||u_i|| \cdot ||e_j|| = \sum_{i=2}^{n} \frac{|\lambda_i(G)|^k}{d^k} \ .$$

Moreover, as G is not a bipartite graph, the property mentioned above implies that for any $2 \leq i \leq n$ we have $|\lambda_i(G)| \leq \max\{|\lambda_2(G)|, |\lambda_n(G)|\} = \lambda(G)$. Hence we have

$$\left| p_j - \frac{1}{n} \right| \leq (n-1) \cdot \left(\frac{\lambda(G)}{d} \right)^k \ ,$$

therefore the fact $0 \leq \lambda(G) < d$ implies that $\lim_{k \to \infty} p_j = \frac{1}{n}$. Summarizing, it follows that the limit distribution of this random walk is the uniform distribution over the vertex set of G, and the evaluation above shows that the speed of convergence to the uniform distribution becomes more rapid as $\lambda(G)$ becomes smaller (that is, the spectral gap becomes larger).

As seen above, graphs with smaller $\lambda(G)$ yield random walks with more rapid convergence to the limit distribution. From such a viewpoint, Ramanujan graphs are well studied from not only purely graph-theoretic motivations but also the importance in applications. For example, it is known that Ramanujan graphs can be constructed by using a certain special kind of elliptic curves (supersingular elliptic curves) and some maps between them called isogenies, and the random walks on those Ramanujan graphs are applied to constructions of cryptographic hash functions. See [1] for the details.

Exercises

Problem 1. For the graph in Figure 7, write down the Laplacian matrix, and verify by using Theorem 8.1 that the graph has 3 spanning trees in total. Moreover, write down all the spanning trees of the graph.

Problem 2. Calculate the number of spanning trees in a complete bipartite graph $K_{n,m}$.

References

[1] D. X. Charles, K. E. Lauter, E. Z. Goren, "Cryptographic Hash Functions from Expander Graphs", Journal of Cryptology, vol.22, pp.93–113, 2009

[2] R. Diestel, "Graph Theorey" (fifth edition), Springer GTM vol.173, Springer, 2017

[3] K. Kimoto, 『レクチャー離散数学』 (in Japanese), SAIENSU-SHA Co., Ltd., 2019

[4] K. Kunen, "Set Theory" (revised edition), College Publications, 2011

[5] R. P. Stanley, "Enumerative Combinatorics", Volume 1, Cambridge University Press, 1997

[6] R. P. Stanley, "Enumerative Combinatorics", Volume 2, Cambridge University Press, 1999