Chapter 5

Topic IV: Some Classical Combinatorics

The purpose of this TOPIC is to give a brief presentation of several classical subjects in the field of combinatorics. The subjects are:

A. Generating functions
B. Inclusion-exclusion
C. Möbius inversion
D. Network flows

The last subject has an extensive literature and involves many interesting algorithmic and data structure related ideas. A thorough discussion of network flows is beyond the scope of this book. We present only the basic ideas. The other subjects represent interesting and important ideas in combinatorics but are of less direct interest to the study of algorithms than the material we have been considering thus far. Generating functions are often useful in obtaining asymptotic complexity results in the theory of algorithms. Inclusion-exclusion and Möbius inversion are interesting classical techniques in enumerative combinatorics where "enumeration" means counting and not listing. These four topics are presented independently of each other. In each case we present only the basic ideas involved. The reader interested in pursuing these subjects further will find appropriate references at the end of Part I. We shall use certain aspects of these topics to motivate some of the material in Part II, and will at certain points refer the reader to the relevant sections within Part II.

5A. GENERATING FUNCTIONS

5A.1 DEFINITION.

The *ordinary generating function* of a sequence $a_0,a_1,\ldots$ is the "formal power series" $a(x) = \sum_{n=0}^{\infty} a_n x^n$.

It is convenient initially to think of the $a_i$ as real or complex numbers but in general they may be polynomials or other functions of $x$, or may themselves be formal power series. In some instances it is important to consider the question
of convergence of the generating function $a(x)$. We shall give some elementary examples of the uses of ordinary generating functions. The references at the end of the chapter contain many additional examples.

### 5A.2 PRODUCTS OF ORDINARY GENERATING FUNCTIONS.

Let $a(x) = \sum_{n=0}^{\infty} a_n x^n$ and $b(x) = \sum_{n=0}^{\infty} b_n x^n$. Then $c(x) = a(x)b(x) = \sum_{n=0}^{\infty} c_n x^n$

where $c_n = \sum_{k=0}^{n} a_k b_{n-k}$. The sequence $(c_0, c_1, \ldots)$ is said to be the convolution of the sequences $(a_0, a_1, \ldots)$ and $(b_0, b_1, \ldots)$.

For example, suppose we wish to find the coefficient of $x^{37}$ in

$$f(x) = \frac{1 - 3x^2 + 4x^7 + 12x^{21} - 5x^{45}}{1 - x}.$$ 

We note that $\frac{1}{1 - x} = \sum_{n=0}^{\infty} x^n = b(x)$ is the generating function of $(b_0, b_1, \ldots) = (1, 1, \ldots)$. If we set $a(x) = \sum_{n=0}^{\infty} a_n x^n = 1 - 3x^2 + 4x^7 + 12x^{21} - 5x^{45}$

then the coefficient of $x^{37}$ in $a(x)b(x)$ is $\sum_{k=0}^{37} a_k b_{37-k} = \sum_{k=0}^{37} a_k = 1 - 3 + 4 + 12 = 14$.

As a variation on the interpretation of products of generating functions consider

$$(1 + x)^m = \sum_{k=0}^{m} \binom{m}{k} x^k$$

the generating function of the sequence $\left( \binom{m}{0}, \binom{m}{1}, \ldots, \binom{m}{m}, 0, 0, \ldots \right)$. Thus $(1 + x)^m (1 + x^{-1})^m$ has $\sum_{k=0}^{m} \binom{m}{k}^2$ as constant term. But $(1 + x)^m (1 + x^{-1})^m = x^{-m}(1 + x)^{2m} = x^{-m} \sum_{k=0}^{2m} \binom{2m}{k} x^k$

which has constant term $\binom{2m}{m}$. Thus we have proved that $\sum_{k=0}^{m} \binom{m}{k}^2 = \binom{2m}{m}$. 


5A.3 EXERCISE.

(1) Find the following indicated coefficients:
   (a) The coefficient of \( x^{20} \) in \( (x^4 + x^5 + x^6 + x^7 + x^8)^3 \).
   (b) The coefficient of \( x^{12} \) in \( \frac{x + 3}{x^2 - 3x + 2} \).
   (c) The coefficient of \( x^{12} \) in \( (1 + x^4)^{-3} \).
   (d) The coefficient of \( x^k \) in \( \sum_{n=0}^{\infty} t^n x^n \).
   (e) The coefficient of \( x^k \) in \( (1 - x)^{-n} \).

(2) Prove that \( \prod_{i=1}^{\infty} (1 + x^i) = \prod_{i=1}^{\infty} (1 - x^{2i-1})^{-1} \) and find the first few coefficients in the power series expansion. (Hint: Write out the first five or six factors in each product and the idea of the proof will become obvious).

As the above examples indicate, it is often important to find "closed form" expressions for generating functions. An important general method is represented in EXAMPLE 5A.4.

5A.4 EXAMPLE: TWO-TERM LINEAR RECURRENCES.

Consider infinite sequences \((a_0,a_1, \ldots)\) which satisfy the "linear two-term recurrence relation" \(a_{k+2} = pa_k + qa_{k+1}\) where \(p\) and \(q\) are constants (all numbers are complex numbers which, of course, includes real numbers). Such a sequence, or its corresponding ordinary generating function \(a(x)\), is determined completely by the first two terms, \(a_0\) and \(a_1\). Thus \(a(x)\) may be specified by giving a pair of complex numbers \((a_0,a_1)\). If \(a(x)\) is determined by \((a_0,a_1)\), and \(b(x)\) is determined by \((b_0,b_1)\), then \(a(x) + b(x)\) is determined by \((a_0 + b_0, a_1 + b_1)\), and \(ra(x)\), where \(r\) is a complex number, is determined by \((ra_0,ra_1)\). Thus, as a vector space, the space of all formal power series that satisfy the given two-term recursion is isomorphic to \(C^2\), the two-dimensional vector space of all complex numbers. The standard "unit vectors" for \(C^2\) are \(i = (1,0)\) and \(j = (0,1)\) with corresponding sequences \((1,0,p,q,p, \ldots)\) and \((0,1,q,p+q^2, \ldots)\) and corresponding power series \(i(x) = 1 + px^2 + qpx^3 + \ldots\) and \(j(x) = x + qx^2 + (p+q^2)x^3 + \ldots\). Any power series whose coefficients satisfy the given two-term recursion can be written as a linear combination of \(i(x)\) and \(j(x)\), but generally this is not much help as \(i(x)\) and \(j(x)\) are awkward expressions. Note, however, that if \(r\) is a root of the equation \(x^2 = qx + p\) then the sequence \((1,r,r^2,r^3, \ldots)\) satisfies the two-term recursion. The corresponding generating function is \(u(x) = 1 + rx + r^2x^2 + \ldots = (1-rx)^{-1}\). If \(s \neq r\) is another root, then its generating function \(v(x) = (1-sx)^{-1}\) is linearly independent from the generating function corresponding to the root \(r\), and hence any generating function that satisfies the given recursion can be written as a linear combination of these two simple generating functions.
Consider, for example, the sequence \((1, 0, -2, -6, -14, \ldots)\) or the generating function \(a(x) = 1 - 2x^2 - 6x^3 - 14x^4 \ldots\) which correspond to \(q = 3\) and \(p = -2\). The equation \(x^2 = 3x - 2\) has two roots, \(r = 1\) and \(s = 2\). Thus the two basis generating functions \(u(x) = (1-x)^{-1}\) and \(v(x) = (1-2x)^{-1}\) may be used to express this generating function. As \((1,0) = 2(1,1) + (-1)(1,2)\), we have \(a(x) = 2(1-x)^{-1} + (-1)(1-2x)^{-1}\). It is immediate from this expression that the \(n^{th}\) coefficient of \(a(x)\), \(a_n\), is given by the formula \(a_n = 2 - 2^n\). An expression such as this is sometimes called a "closed form" expression for \(a_n\).

5A.5 EXERCISE.

1. A Fibonacci sequence is a sequence \((a_0, a_1, a_2, \ldots)\) that satisfies the linear two-term recurrence \(a_{k+2} = a_k + a_{k+1}\) (i.e., \(p = q = 1\) in EXAMPLE 5A.4). Find the basis series \(u(x)\) and \(v(x)\) for these sequences (as in EXAMPLE 5A.4). Find a closed form expression for the \(n^{th}\) term of the Fibonacci sequence \((0, 1, 1, 2, 3, 5, 8, \ldots)\).

2. Given that the sequence \((0, 1, 3, 13, 51, \ldots)\) satisfies a linear two-term recurrence, find the recurrence relation and the basis series \(u(x)\) and \(v(x)\). Find a closed form expression for the \(n^{th}\) term.

3. How should the results of EXAMPLE 5A.4 be modified if \(r = s\)? (Hint: If \(u(x) = \frac{1}{1-rx}\), take \(v(x) = \frac{d}{dr} \left( \frac{1}{1-rx} \right)\).)

4. Suppose that the sequence \((a_0, a_1, a_2, \ldots)\) satisfies a recurrence relation of the form \(a_{k+2} = c + pa_k + qa_{k+1}\) where \(c\) is a fixed complex number. How should EXAMPLE 5A.4 be modified?

5. Suppose that the sequence \((a_0, a_1, a_2, \ldots)\) satisfies a linear three-term recurrence \(a_{k+3} = pa_k + qa_{k+1} + ta_{k+2}\). How should the results of EXAMPLE 5A.4 and EXERCISE 5A.5 (3) above be modified? Give some examples of the various possibilities. The theory of the linear m-term recurrence follows in the obvious way but becomes increasingly intractable in practice.

5A.6 EXAMPLE: GENERATING FUNCTIONS FOR COMPOSITIONS.

In TABLE 3.25 and in the associated discussion we considered nondecreasing functions. Consider, for example, the nondecreasing functions from \(4\) to \(3\). In one line notation, \((1, 1, 1, 1), (1, 1, 1, 2), (1, 2, 2, 2), (1, 1, 2, 2),\) and \((1, 1, 2, 3)\) are such functions. Another way to "code" or represent these functions is to specify instead triples \((x_1, x_2, x_3)\) where \(x_1\) is the number of \(1's\), \(x_2\) is the number of \(2's\), and \(x_3\) is the number of \(3's\) in the corresponding nondecreasing function. The above functions would be represented by \((4, 0, 0), (3, 1, 0), (1, 3, 0), (2, 2, 0),\) and \((2, 1, 1)\), respectively. The sum of the entries in these triples must, of course,
add up to 4. A composition of \( n \) into \( p \) parts is a sequence \((x_1, x_2, \ldots, x_p)\) of non-negative integers whose sum is \( n \). As we have just seen, there is a natural bijection between the compositions of \( n \) into \( p \) parts and the nondecreasing functions of \( n \) to \( p \). In connection with TABLE 3.25, we showed that there are
\[
\binom{n + p - 1}{n}
\]
such nondecreasing functions. In terms of generating functions, let
\[
a(x) = a_0 + a_1x + \ldots
\]
be the generating function of the sequence \((a_0, a_1, \ldots)\)
where \( a_n \) is the number of compositions of \( n \) into \( p \) parts. If \( b(x) = 1 + x + x^2 + \ldots \) is the generating function of the sequence \((b_0, b_1, b_2, \ldots)\) where \( b_i = 1 \) for all \( i \), then \( b(x) \) can be thought of as the generating function for the number of compositions of \( n \) into one part. Clearly, \( c(x) = a(x)b(x) \) is the generating function of the compositions of \( n \) into \( p + 1 \) parts. Thus, inductively, we see that \( a(x) = (b(x))^p = (1 - x)^{-p} \) is the generating function for the number of compositions of \( n \) into \( p \) parts. By the binomial theorem, \( (b(x))^p = \sum_{n=0}^{\infty} \binom{-p}{n} (-x)^n \). But,
\[
\binom{-p}{n} = \frac{-p(-p-1)(-p-2)\ldots(-p-n+1)}{n!}
= (-1)^n \binom{n+p-1}{n}
\]
and hence we derive again the formula for the number of nondecreasing functions or, equivalently, the number of compositions of \( n \) into \( p \) parts. By modifying this argument slightly, we can put various restrictions on the compositions. If no part is zero in the composition of \( n \) into \( p \) parts, then the generating function becomes \((c(x))^p = (x + x^2 + \ldots)^p \) or \( \left(\frac{x}{1-x}\right)^p \). The generating function \( f(x) = \sum_{p=0}^{\infty} (c(x))^p \) is the generating function of the number of compositions of \( n \) without zero parts. Clearly,
\[
f(x) = \frac{1}{1 - c(x)} = \frac{1 - x}{1 - 2x} = 1 + \sum_{n=1}^{\infty} 2^{n-1}x^n.
\]

One can easily derive this result directly by associating compositions of \( n \) with functions \( 2^{n-1} \). The generating function for the number of compositions of \( n \) into 10 parts where the first part can be at most 5 and the last part can never be zero is \((1 + x + \ldots + x^5)(1-x)^{-8}(x/(1-x))\). As the references at the end of the chapter testify, an endless variety of cute word problems can be made up based on the idea of compositions with restricted parts.

5A.7 EXERCISE.

(1) Relate the following two problems to the discussion of EXAMPLE 5A.6:
(a) How many ways are there to get a sum of 36 when 12 dice are rolled?
(b) Suppose there are four large boxes. The first box is filled with green balls, the second is filled with blue balls, the third is filled with red balls, and the fourth is filled with yellow balls. How many ways are there to select 24 balls from these boxes if at least one ball must be selected from each box?

(2) Give an expression for the generating function for the number of compositions of \( n \) into \( p \) parts if the \( k^{th} \) part is divisible by \( k \), \( k = 1, \ldots, p \).

5A.8 EXAMPLE: GENERATING FUNCTIONS FOR INTEGRAL PARTITIONS.

Consider a sequence \( 1 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \lambda_4 \leq 4 \) whose sum \( \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 9 \). For example, \((1,2,2,4)\) and \((2,2,2,3)\) are two such sequences. Such sequences are "integral partitions of 9 into four parts with maximum part size 4". We have considered integral partitions in the paragraph just following EXERCISE 3.50, in EXERCISE 3.51(2), and in FIGURE 4.69 and EXERCISE 4.70. The sequence \((1,4,4)\) is also an integral partition of 9 with maximum part size 4 but it has only three parts. An integral partition such as these may be described by the notation \( 1^{j_1} 2^{j_2} 3^{j_3} 4^{j_4} \) which is read "\( j_1 \) 1's, \( j_2 \) 2's, \( j_3 \) 3's, \( j_4 \) 4's." We must have \( j_1 + 2j_2 + 3j_3 + 4j_4 = 9 \). The number of parts is \( j_1 + j_2 + j_3 + j_4 \). With this notation \((1,2,2,4)\) would be written \( 1^1 2^2 4^1 \) and \((2,2,2,3)\) would be written \( 2^3 3^1 \). Note that we omit terms such as \( 3^0 \) for convenience of notation. We could also describe such integral partitions by expressions of the form \( (x)_{j_1} (x^2)_{j_2} (x^3)_{j_3} (x^4)_{j_4} = x^9 \). From this latter description, it is easy to see that the number of integral partitions of 9 with maximum part size 4 is the coefficient of \( x^9 \) in the expansion of the product \( (1 + x + x^2 + \ldots) (1 + (x^2) + (x^2)^2 + \ldots) (1 + (x^3) + (x^3)^2 + \ldots) (1 + (x^4) + (x^4)^2 + \ldots) \). The reader should note that each integral partition of 9 with maximum part size 4 corresponds to a composition \((j_1,2j_2,3j_3,4j_4)\) of 9 into four parts where the \( k^{th} \) part is divisible by \( k \) (see EXERCISE 5A.7(2)). In any case, the generating function for the number of integral partitions of \( n \) with maximum part size 4 can obviously be written as the product \( (1 - x)^{-1} (1 - x^2)^{-1} (1 - x^3)^{-1} (1 - x^4)^{-1} \). More generally, we have that the generating function for the number of integral partitions of \( n \) with maximum part size \( p \) is \( \prod_{i=1}^{p} (1 - x^i)^{-1} \) and the generating function for the number of integral partitions of \( n \) (no restriction on part size) is \( \prod_{i=1}^{\infty} (1 - x^i)^{-1} \). Many variations on these ideas are possible.

5A.9 EXERCISE.

(1) Let \( c_n \) denote the number of ways to represent \( n \) cents in terms of pennies, nickels, dimes, and quarters. Find an expression for the generating function

\[
 c(x) = \sum_{n=1}^{\infty} c_n x^n \text{ and evaluate } c_{100}.
\]
(2) Explain why the generating function for the number of integral partitions with no repeated parts is given by \( \prod_{i=1}^{\infty} (1 + x^{i}) \) and the generating function for the number of integral partitions with each part odd is given by \( \prod_{i=1}^{\infty} (1 - x^{2i-1})^{-1} \). Note by EXERCISE 5A.3(2) we have the remarkable fact that these two classes of integral partitions are "equinumerous" as their generating functions are equal (see Remmel, 1982, in Part I References).

(3) Let \((a_n)\) denote the sequence \((a_0, a_1, \ldots)\) with generating function \(a(x)\). Note that \((d/dx)a(x)\) is the generating function for the sequence \(((n + 1)a_{n+1})\) and \(x(d/dx)a(x)\) is the generating function for the sequence \((na_n)\). We have already noticed that \((1-x)^{-1}a(x)\) is the generating function for the sequence \((a_0 + a_1 + \ldots + a_n)\). By combining these types of operations, generating functions of more complicated sequences can be constructed. Thus, \(b(x) = (1-x)^{-1}x(d/dx)a(x)\) is the generating function of the sequence \((a_1 + 2a_2 + \ldots + na_n)\). For example, if we take \(a(x) = (1-x)^{-1}\) then \(b(x) = x/(1-x)^3\) must have as its coefficient of \(x^n\) the sum \(1 + 2 + \ldots + n\). But, using EXERCISE 5A.3(1(e)), we see that the coefficient of \(x^n\) in \((1-x)^{-3}\) is \(\binom{-3}{n}(-1)^n\) and hence the coefficient of \(x^n\) in \(b(x)\) is \(\binom{n+1}{2}\). This proves the standard result (easily proved geometrically or by induction) that \(1 + 2 + \ldots + n = \binom{n+1}{2}\). Using these techniques, derive some more complicated identities of your own. One obvious one to try is \(1^2 + 2^2 + \ldots + n^2 = \binom{n+1}{3} + \binom{n+2}{3}\).

(4) A product of three matrices \(A_1A_2A_3\) can be taken in two ways: \((A_1A_2)A_3\) and \(A_1(A_2A_3)\). In the first case, \(A_1\) and \(A_2\) are multiplied first, and in the second case \(A_2\) and \(A_3\) are multiplied first. For four matrices, there are five ways to form the product: \((A_1A_2)(A_3A_4), (A_1A_2)(A_3)A_4, (A_1)(A_2A_3)A_4, A_1(A_2)(A_3A_4), \) and \(A_1(A_2A_3A_4)\) (all yield the same answer).

(a) Let \(p_n\) denote the number of ways to take the product of \(n\) matrices.

Show that \(p_n = p_1p_{n-1} + p_2p_{n-2} + \ldots + p_{n-1}p_1\) when \(n > 1\).

(b) Let \(p(x) = p_1x + p_2x^2 + \ldots + p_nx^n + \ldots\) be the generating function for the \(p_n\). Show that \(p(x) - x = (p(x))^2\).

(c) Using the quadratic formula, solve the equation of part (b) for \(p(x)\) to obtain \(p(x) = (1 - (1 - 4x)^{1/2})/2\).

(d) Using the binomial theorem \((1 + z)^a = \sum_{n=0}^{\infty} \binom{a}{n} z^n\) with \(a = 1/2\) and \(z = -4x\), show that \(p_n = \frac{1}{n} \binom{2n-2}{n-1}\). The \(p_n\) are called "Catalan" numbers.
For certain computations the ordinary generating function becomes awkward, and it is best to introduce slight variations on this idea.

5A.10 DEFINITION.
The exponential generating function of a sequence \((a_0, a_1, \ldots)\) is the formal power series \(\sum_{n=0}^{\infty} a_n x^n/n!\).

5A.11 PRODUCTS OF EXPONENTIAL GENERATING FUNCTIONS.
Let \(a(x) = \sum_{n=0}^{\infty} a_n x^n/n!\) and \(b(x) = \sum_{n=0}^{\infty} b_n x^n/n!\) and \(c(x) = a(x)b(x)\)

\[= \sum_{n=0}^{\infty} c_n x^n/n!\]

Then \(c_n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array}\right) a_k b_{n-k}\).

Note that the exponential generating function of the sequence \((1, 1, \ldots)\) is \(e^x\). We shall give some examples of the use of exponential generating functions. Many additional examples can be found in the references.

5A.12 EXAMPLE: EXPONENTIAL GENERATING FUNCTIONS AND DERANGEMENTS.
We use the standard notation \(S_n\) for the set of all permutations of \(n\). For \(Q \subseteq n\), let \(D(Q) = \{f: f \in S_n \text{ and } f(i) = i \text{ if and only if } i \in n - Q\}\). In other words, \(f\) is in \(D(Q)\) if the set of elements of \(n\) that are fixed by \(f\) is \(n - Q\). The set \(D(Q)\) will be called the set of \(Q\)-derangements of \(n\). The set of permutations \(D(n)\) is called the set of derangements of \(n\). Let \(d_n = |D(n)|\) denote the number of derangements of \(n\) \((d_0 = 1)\). Clearly, if \(|Q| = q\) then \(|D(Q)| = d_q\). Obviously, \(S_n\) is the disjoint union of the \(D(Q)\) as \(Q\) ranges over all subsets of \(n\). Thus we have \(n! = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array}\right) d_k\). Let \(d(x) = \sum_{n=0}^{\infty} d_n x^n/n!\) be the exponential generating function of the sequence \((d_n)\). Using the rule for PRODUCTS OF EXPONENTIAL GENERATING FUNCTIONS 5A.11, we must have \(\sum_{n=0}^{\infty} n! x^n/n! = d(x)e^x\) or \(d(x) = e^{-x}(1 - x)^{-1} = \sum_{n=0}^{\infty} c_n x^n/n!\) where \(c_n = n! \sum_{k=0}^{n} (-1)^k/k!\). For large \(n\), \(c_n\) is thus approximately \(n!/e\) \((e\) is about 2.7). This is the basis for the assertion that, if an irate hatcheck person scrambles the hatcheck receipts, the probability that no one gets their own hat back is \(e^{-1}\). We shall derive this result from another point of view in the section on INCLUSION-EXCLUSION (EXAMPLE 5B.3).
EXERCISE.

(1) Use generating functions to derive the recurrence relation \( d_n = nd_{n-1} + (-1)^n. \) Hint: \( d(x)(1-x) = e^{-x}. \)

(2) Use generating functions to derive the recurrence relation \( d_{n+1} = n(d_n + d_{n-1}). \) Explain why this recurrence relation is true using only the definition of derangements. (See EXERCISE 3.53.) Show that \( x d(x) + x d'(x) = d'(x). \)

(3) Consider \( e^x(1-x)^{-p} = \sum_{n=1}^{\infty} c_n x^n / n!. \) It turns out that \( c_n \) is the number of ways to choose \( k \) balls (\( k = 0, 1, \ldots, n \)) from \( n \) balls, labeled \( 1, \ldots, n \), and distribute them into \( p \) boxes where the order of the balls in the boxes matters. Thus \( [1, 2, 3] \) is a different distribution of three balls into two boxes than \( [2, 1, 3] \). Explain why the \( c_n \) have this interpretation.

We conclude our discussion of the exponential generating function with a discussion of "polynomials of binomial type." There are a number of approaches to the problem of giving a combinatorial model for generating functions. The reader will find some suggested articles in the references at the end of Part I.

We now give an elementary version of such a model. Imagine that, given a set \( S \) of integers, we are able to print the integers in \( g_1 \) different colors where \( s \) = \( |S| \) (all integers in the set are printed the same color, but we have \( g_i \) different colors to choose from). We might denote by \( S_{\text{red}} \) the set \( S \) with all integers printed red, \( S_{\text{blue}} \) all integers printed blue, etc. Our basic assumption is that the number of colors that the set can be printed in depends only on the size of the set. Consider an ordered partition \( \{A_1, A_2, \ldots, A_p\} \) of \( n \) with multinomial index \( (a_1, a_2, \ldots, a_p) \). Such partitions were discussed in connection with FIGURE 3.3, EXERCISE 3.4, in the paragraph following EXERCISE 3.27, and in connection with FIGURE 3.28 and EXERCISE 3.29. We only need to recall here that the \( A_i \) form a pair-wise disjoint collection of subsets of \( n \) whose union is \( n \). For each \( i, a_i = |A_i| \). We assume that the \( a_i \) > 0. For each set \( A_i \), the integers of \( A_i \) can be printed in \( g_{a_i} \) different choices of colors (for each choice, all elements of \( A_i \) are printed in the same color as above). This means that there are \( g_{a_1} g_{a_2} \cdots g_{a_p} \) different ways to print the given ordered partition. There are \( n! / a_1! a_2! \cdots a_p! \) ways to select an ordered partition on \( n \) with multinomial index \( (a_1, a_2, \ldots, a_p) \). This number is called the multinomial coefficient. (See EXERCISE 5A.34(1).)

NOTATION FOR THE MULTINOMIAL COEFFICIENT

\[
\frac{n!}{a_1! a_2! \cdots a_p!} = \binom{n}{a_1, a_2, \ldots, a_p}.
\]

Each set \( A_i \) in an ordered partition is called a block of the partition. The ordered partition \( \{A_1, A_2, \ldots, A_p\} \) has \( p \) blocks (the blocks are nonempty in this
discussion). The total number of color printed ordered partitions with \( p \) blocks is given by IDENTITY 5A.15.

5A.15 IDENTIFY: THE NUMBER \( c_n^*(g,p) \) OF COLOR PRINTED ORDERED PARTITIONS OF \( n \) WITH \( p \) BLOCKS.

\[
c_n^*(g,p) = \sum_{a_1 + \ldots + a_p = n} g_{a_1} g_{a_2} \ldots g_{a_p} \begin{pmatrix} n \\ a_1 a_2 \ldots a_p \end{pmatrix}.
\]

The sum is over all compositions \((a_1, a_2, \ldots, a_p)\) of \( n \), where \( a_i > 0 \) for all \( i \).

We now observe that each of the \( p! \) different rearrangements of a color printed ordered partition with \( p \) blocks is distinct. Thus, if we divide \( c_n^*(g,p) \) by \( p! \) then we obtain \( c_n(g,p) \), the total number of color printed set (unordered) partitions of \( n \) into \( p \) blocks. This modification of IDENTITY 5A.15 is given by IDENTITY 5A.16.

5A.16 IDENTIFY: THE NUMBER, \( c_n(g,p) \), OF COLOR PRINTED SET PARTITIONS OF \( n \) WITH \( p \) BLOCKS.

\[
c_n(g,p) = \frac{1}{p!} \sum_{a_1 + \ldots + a_p = n} g_{a_1} g_{a_2} \ldots g_{a_p} \begin{pmatrix} n \\ a_1 a_2 \ldots a_p \end{pmatrix}.
\]

The sum is over all compositions \((a_1, a_2, \ldots, a_p)\) of \( n \), where \( a_i > 0 \) for all \( i \).

In EXAMPLE 5A.17 we discuss the relationship between exponential generating functions and color printed partitions. This model relates to a number of basic combinatorial problems.

5A.17 EXAMPLE: COLOR PRINTED PARTITIONS AND EXPONENTIAL GENERATING FUNCTIONS.

Consider \( e^{xg(u)} \) where \( g(u) = g_1 u + g_2 u^2/2! + g_3 u^3/3! + \ldots \) is the exponential generating function of the sequence \((g_1, g_2, \ldots)\). Using the power series expansion for the exponential function, we write \( e^{xg(u)} = 1 + (xg(u)) + (xg(u))^2/2! + \ldots \). This expression can be rewritten in the form \( 1 + p_1(x)u + p_2(x)u^2/2! + \ldots \). This latter expression may be regarded as the exponential generating function of a sequence of polynomials \( \{p_n(x)\} \). In order to avoid convergence problems, we have assumed that the constant term of \( g(u) \) is zero. We wish to express the coefficients of the polynomials \( p_n(x) \) in terms of the \( g_i \). Consider a typical term \((xg(u))^{p}/p! \). The coefficient of \( x^p \) in \( p_n(x) \) will be

\[
\frac{n!}{p!} \sum \frac{g_{a_1} g_{a_2} \ldots g_{a_p}}{a_1! a_2! \ldots a_p!}.
\]
where the sum is over all compositions \((a_1, a_2, \ldots, a_p)\) of \(n\) where \(a_i > 0\) for all \(i\). This sum is obviously \(c_n(g,p)\) of IDENTIFY 5A.16. Thus we have derived IDENTIFY 5A.18.

**5A.18 IDENTIFY: COEFFICIENTS OF \(p_n(x)\) ARE THE \(c_n(g,p)\) OF IDENTIFY 5A.16.**

\[
e^{xg(u)} = 1 + p_1(x)u + \ldots + p_n(x)u^n/n! + \ldots
\]

where \(p_n(x) = \sum_{p=1}^{n} c_n(g,p)x^p\).

We assume that \(g(u) = g_1u + g_2u^2/2! + \ldots\). Henceforth we shall also assume that \(g_1 \neq 0\).

There are many interesting combinatorial interpretations of IDENTIFY 5A.18. We have already discussed color printed partitions. For these we assume that for the integers in a set \(S \subseteq n\) with \(|S| = s\), we have exactly \(g_s\) choices for the color of the print (all integers are printed the same color once a choice is made). We showed that a partition of \(n\) into \(p\) blocks can be printed in \(c_n(g,p)\) different ways, given by IDENTIFY 5A.16. The model of "color printed partitions" may seem too restrictive to be of much interest. However, if we let our imaginations soar a bit, we may interpret the idea of coloring a set more generally as imposing some structure on the set. For example, given a set \(S \subseteq n\) with \(|S| = s\), there are \(s!\) ways to linearly order the elements of \(S\). We might think of such rearrangement as a "color." In this case \(g_s = s!\). A "color printed partition with \(p\) blocks" thus corresponds to a partitioning of a set of \(n\) elements into \(p\) blocks and the selection of a linear order on each block.

As another example, if \(|S| = s\) there are \((s-1)!\) ways to arrange the elements of \(S\) in cyclic order. In this case, \(g_s = (s-1)!\) and a color printed partition of \(n\) with \(p\) blocks corresponds to a partitioning of \(n\) into \(p\) blocks and the selection of a cyclic order on the elements of each block. The reader familiar with the disjoint cycle notation for permutations (EXAMPLE 4.34, EXERCISE 4.35) will recognize that, in this case, \(c_n(g,p)\) is the number of permutations on \(n\) with exactly \(p\) cycles. Looking ahead to CHAPTER 6, we may let \(g_s\) denote the number of connected graphs with \(s\) vertices (CHAPTER 6, DEFINITION 6.11). In this case, \(c_n(g,p)\) is the number of graphs with \(n\) vertices and \(p\) components. Or, let \(g_s\) be the number of trees with \(s\) vertices (CHAPTER 6, DEFINITION 6.12.). In this case \(c_n(g,p)\) is the number of forests with \(n\) vertices and \(p\) components. There are many other possibilities. One that we have not yet mentioned is the trivial case where \(g_s = 1\) for all \(s\). In this case, \(c_n(g,p)\) corresponds to the number of partitions of \(n\) with \(p\) blocks (Stirling numbers of the second kind, TABLE 3.49).
5A.19  EXAMPLE: SET PARTITIONS.

As mentioned above, if \( g_s = 1 \) for all \( s \), then \( c_n(g,p) \) is the number of partitions of \( n \) with \( p \) blocks. In this case, \( g(u) = u + u^2/2! + \ldots = e^u - 1 \) and \( e^{xg(u)} \) is \( e^{x(u-1)} \). If we set \( x = 1 \) we obtain \( e^{u-1} \) as the generating function for the number of set partitions of a set of \( n \) elements (the coefficient of \( u^n/n! \)).

5A.20  EXAMPLE: LAH NUMBERS.

If \( g_s = s! \) then \( c_n(g,p) \) is the number of partitions of \( n \) with \( p \) nonempty ordered blocks. We might think of constructing such a partition by first specifying a permutation of \( n \). For example, let \( n = 6 \) and consider \( 6\ 3\ 5\ 2\ 4\ 1 \). Next we choose \( p - 1 \) of the \( n - 1 \) spaces between symbols in this permutation, and insert vertical bars in these spaces to define the \( p \) ordered blocks. In our example, if \( p = 3 \) we might choose \( 6\ 3\ 5\ \vert\ 2\ 4\ \vert\ 1 \), giving the ordered blocks \((6,3,5), (2,4), (1)\). This process can be carried out in \( n! \binom{n-1}{p-1} \) ways. In this way we would construct all ordered partitions of \( n \) into \( p \) nonempty ordered blocks. To get unordered partitions into \( p \) ordered blocks we divide by \( p! \) giving \( \frac{n!}{p!} \binom{n-1}{p-1} \), the so called “Lah numbers.” But, \( g(u) = u + u^2 + u^3 = \ldots = u/(1-u) \). From IDENTITY 5A.18 we thus have \( e^{xu/(1-u)} = 1 + p_1(x)u + p_2(x)u^2/2! + \ldots \) where \( p_n(x) = \sum_{p=1}^{n} \frac{n!}{p!} \binom{n-1}{p-1} x^p \).

5A.21  EXAMPLE: STIRLING NUMBERS OF THE FIRST KIND.

A permutation on \( n \) with only one cycle (including cycles of length one) is called a “full cycle” on \( n \). For example, \((1, 4, 5, 3, 2, 6)\) is a full cycle on \( 6 \). A full cycle on \( 6 \) can always start with \((1, \ldots)\) and the remaining entries can be any of the \( 5! \) permutations of \( 2, \ldots, 6 \). All full cycles on \( 6 \) can be constructed in this way. Thus by defining \( g_s = (s-1)! \) we are counting full cycles on a set of size \( s \). For this choice of \( g_s \), the \( c_n(g,p) \) count the number of permutations of \( n \) with \( p \) cycles. These numbers are called the “Stirling numbers of the first kind” (strictly speaking, the absolute values of these numbers). For this choice of \( g_s \), \( g(u) = u + u^2/2 + u^3/3 + \ldots = -\ln(1-u) \). Again, by IDENTITY 5A.18, \( e^{-\ln(1-u)} = 1 + p_1(x)u + p_2(x)u^2/2! + \ldots \) is the exponential generating function for these polynomials. It can be shown that \( p_n(x) = (x)^n \), the “upper factorial” polynomial, where \( (x)^n = x(x+1) \ldots (x+n-1) \).

5A.22  EXAMPLE: GRAPHICAL STRUCTURES.

Many types of graphical structures are constructed from “connected” structures of the same type. A graph has its connected components (DEFINITION 6.11 of CHAPTER 6). A forest is a graph whose components are trees
(DEFINITION 6.12 of CHAPTER 6). A rooted forest is a graph whose components are rooted trees (DEFINITION 6.23 of CHAPTER 6). In all of these cases, if $g_s$ is the number of connected structures with vertex set of size $s$ then $c_n(g,p)$ is the number of structures with vertex set of size $n$ and $p$ components. An interesting example is the case of rooted forests. We shall see in CHAPTER 6 that $g_s = s^{s-1}$ (LEMMA 6.20, EXERCISE 6.22(1)). We shall see in the following material that, for this case, $p_n(x) = x(x+n)^{n-1}$ (EXAMPLES 5A.32, 5A.33).

We conclude our discussion of the exponential generating function with a brief presentation of some linear algebraic properties of sequences of polynomials. Note that the sequence of polynomials $\{x^n/n!\}$ is a basis for the vector space of all polynomials. The differentiation operator $D$ is a linear transformation on this vector space, and "shifts" this basis in the sense that $D(x^n/n!) = x^{n-1}/(n-1)!$ for all $n \geq 1$.

5A.23 DEFINITION.

A sequence of polynomials $\{p_n(x)\}$ where the degree of $p_n$ is $n$ for all $n$ will be called a basis sequence. If $p_0 = 1$ and $p_n(0) = 0$ for all $n > 0$ then the basis sequence will be called a normalized basis sequence. A linear transformation $S$ on polynomials for which there exists a basis sequence $\{p_n\}$ such that $S(p_n/n!) = p_{n-1}/(n-1)!$ for all $n \geq 1$ and $Sp_0 = 0$ will be called a basis shift operator with shift basis $\{p_n/n!\}$. (Note: To say that the degree of a polynomial is zero implies that the polynomial is a nonzero constant.)

Obviously, a basis sequence is a basis for the vector space of all polynomials. If $\mathcal{P}_n$ denotes the vector space of all polynomials of degree less than or equal to $n$ then $S$ is a linear transformation from $\mathcal{P}_n$ to $\mathcal{P}_{n-1}$ of rank $n-1$ and nullity 1. This observation is the basic fact needed to prove LEMMA 5A.24.

5A.24 LEMMA.

Let $S$ be a basis shift operator. Suppose that both $\{p_n/n!\}$ and $\{q_n/n!\}$ are shift bases for $S$. Then, there exists a sequence of numbers $(a_0, a_1, \ldots, a_n, \ldots)$ such that $a_0 \neq 0$ and for all $n$, $q_n/n! = \sum_{k=0}^{n} a_{n-k} p_k/k!$. Conversely, if such a sequence relates the two sequences $\{p_n/n!\}$ and $\{q_n/n!\}$, then if either sequence is a shift basis for $S$, the other is also.

Proof. The converse is left to the reader. Suppose that $\{p_n/n!\}$ and $\{q_n/n!\}$ are shift bases for $S$. We use induction. As $p_0$ and $q_0$ are both of degree 0 (i.e., nonzero constants) we have $q_0 = a_0 p_0$. Assume that there is a sequence $(a_0, a_1, \ldots, a_n)$ such that for all $m \leq n$, $q_m/m! = \sum_{k=0}^{m} a_{m-k} p_k/k!$. Let $f =$
\[ q_{n+1}/(n+1)! - \sum_{k=1}^{n+1} a_{n+1-k}p_k/k!. \] Then by our hypotheses, \( S(f) = 0. \) Since the basis shift operator \( S \) has nullity 1, and \( p_0 \) is in the null space of \( S, \) we have \( f = a_{n+1}p_0 \) for some number \( a_{n+1} \) (this is how \( a_{n+1} \) is determined). Thus

\[ q_{n+1}/(n+1)! = \sum_{k=0}^{n+1} a_{n+1-k}p_k/k!. \]

**COROLLARY 5A.25** follows from **LEMMA 5A.24.** (EXERCISE 5A.34(2)).

**5A.25 COROLLARY.**

Let \( S \) be a basis shift operator. Then there is one and only one normalized basis sequence \( \{p_n\} \) such that \( \{p_n/n!\} \) is a shift basis for \( S. \)

**COROLLARY 5A.26 is a rephrasing of LEMMA 5A.24.**

**5A.26 COROLLARY.**

Let \( S \) be a basis shift operator with shift basis \( \{p_n/n!\}. \) A basis sequence \( \{q_n/n!\} \) is also a shift basis for \( S \) if and only if there is a formal power series \( f(S) = a_0 + a_1S + a_2S^2 + \ldots \) with \( a_0 \neq 0 \) such that \( f(S)p_n = q_n \) for all \( n. \)

**5A.27 DEFINITION.**

The sequence \( \{p_n\} \) of polynomials defined by **IDENTITY 5A.18** (with the assumption \( g_1 \neq 0 \)) defines a normalized basis sequence which we call the **normalized basis sequence of** \( g(u). \)

**LEMMA 5A.28** is an important technical lemma that relates **LEMMA 5A.24** and **DEFINITION 5A.27.** If \( g(u) = g_1u + g_2u^2 + \ldots \) is a formal power series with \( g_1 \neq 0 \) then there always exists a formal power series \( h(u) = h_1u + h_2u^2 + \ldots \) with \( h_1 \neq 0 \) such that \( g(h(u)) = u. \) Assume all coefficients are real or complex numbers. The power series \( h(u) \) is called the **compositional inverse** of \( g(u) \) and is denoted by \( g^{-1}(u) \) (see EXERCISE 5A.34(3)). If \( D \) denotes differentiation with respect to \( x, \) and if \( f(u) \) is a formal power series with real or complex coefficients, then \( f(D) \) is defined as an operator on polynomials. Similarly, \( D \) or \( f(D) \) acts on formal power series with polynomial coefficients by transforming each coefficient. If \( \sum_{m=0}^{\infty} c_n(x)u^n \) is such a formal power series and \( g(u) \) is as above, we can define a transformation \( A_g \sum_{m=0}^{\infty} c_n(x)u^n = \sum_{m=0}^{\infty} c_n(x)(g(u))^n. \) Thus \( D, f(D), \) and \( A_g \) all act as transformations on power series with polynomial coefficients.
5A.28 LEMMA.

Let \( \{p_n\} \) be the normalized basis sequence (DEFINITION 5A.27) of the formal power series \( g(u) = g_1u + g_2u^2 + \ldots \) and let \( S = g^{-1}(D) \) where \( D \) is differentiation with respect to \( x \). Then \( S \) is a basis shift operator with shift basis \( \{p_n/n!\} \).

**Proof.** As above, let \( A_g \) denote the operator which replaces \( u \) with \( g(u) \) in a formal power series. Clearly, as transformations on power series with polynomial coefficients, \( DA_g = A_gD \) and hence \( f(D)A_g = A_gf(D) \) for any formal power series \( f \). Note that \( g^{-1}(D)e^{xg(u)} = g^{-1}(D)A_g e^{xu} = A_g g^{-1}(D)e^{xu} = A_g g^{-1}(u)e^{xu} = ue^{xg(u)} \). Interpreting the equality between the first and last of these expressions in terms of coefficients of powers of \( u \) gives the result.

Of course, not all normalized basis sequences can be associated with a formal power series \( g(u) \) in the sense of DEFINITION 5A.27. LEMMA 5A.29 is sometimes useful in identifying those that do have this property.

5A.29 LEMMA.

Let \( \{p_n\} \) be a normalized basis sequence and let \( h(u) = h_1u + h_2u^2 + \ldots , \) where \( h_1 \neq 0 \), be a formal power series. If \( \{p_n/n!\} \) is a shift basis for \( h(D) \) then \( p_n \) is the normalized basis sequence of \( g(u) = h^{-1}(u) \).

**Proof.** Let \( \{q_n\} \) be the normalized basis sequence associated with \( g(u) \). By LEMMA 5A.28, \( \{q_n/n!\} \) is a shift basis for \( g^{-1}(D) = h(D) \). By COROLLARY 5A.25, \( p_n = q_n \) for all \( n \).

We shall give several examples of the use of LEMMA 5A.29, but first we define an important class of operators on polynomials that are power series in the differentiation operator.

5A.30 DEFINITION.

For any number \( a, E^a \) is the linear operator on polynomials defined by \( E^a p(x) = p(x+a) \). \( E^a \) is called "translation by \( a \)," or "(argument) shift by \( a \)." We use the former terminology to avoid confusion with the basis shift operators.

**Remark.** By Taylor's theorem, \( p(x+a) = \sum_{k=0}^{\infty} \frac{a^k}{k!} D^k p(x) \) and thus \( E^a = \sum_{k=0}^{\infty} \frac{a^k}{k!} D^k = e^{aD} \) is a formal power series in the differentiation operator \( D \).
5A.31 EXAMPLE: APPLICATIONS TO ABEL AND FACTORIAL POLYNOMIALS.

The polynomials \( (x)_n = x(x-1) \cdot \cdots \cdot (x-n+1) \) and \( (x)^n = x(x+1) \cdot \cdots \cdot (x+n-1) \) are called the lower and upper factorial polynomials respectively. The reader may easily verify that \( E^{-1} = e^D - 1 \) is a basis shift operator with shift basis \( (x)_n/n! \). Similarly, \( 1 - E^{-1} = 1 - e^{-D} \) is a basis shift operator with shift basis \( (x)^n/n! \). Thus \( (x)_n \) is the normalized basis sequence of \( g(u) = h^{-1}(u) \) where \( h(u) = e^u - 1 \). In this case we have \( g(u) = \ln(1 + u) \). Similarly, \( (x)^n \) is the normalized basis sequence for \( g(u) = -\ln(1-u) \). A more interesting example is the sequence of polynomials \( x(x+an)^{n-1} \), called the Abel polynomials. A basis shift operator for these polynomials is \( DE^{-a} = De^{-aD} \) (EXERCISE 5A.34(4)). Thus, the sequence of Abel polynomials is a normalized basis sequence for \( g(u) = h^{-1}(u) \) where \( h(u) = ue^{-au} \). No simple closed form expression for \( g(u) \) is known. The coefficient of \( x \) in \( x(x+an)^{n-1} \) is \( (an)^{n-1} \) thus by IDENTITY 5A.18 and

\[ 5A.16, \quad g(u) = \sum_{n=1}^{\infty} \frac{(an)^{n-1} u^n}{n!}. \]

5A.32 EXAMPLE: ABEL POLYNOMIALS AND ENUMERATION OF TREES.

Suppose that \( \{p_n(x)\} \) is the normalized basis sequence for the power series \( g(u) \). From IDENTITIES 5A.18 and 5A.16, we see that the sequence of polynomials \( \{p_n(x)\} \) is determined completely by the coefficients of \( x \) alone, since these values determine the power series \( g(u) \). Suppose that \( p_n(x) = \sum_{p=1}^{n} c_n(g,p)x^p \) where \( c_n(g,p) \) is the number of rooted forests on \( n \) vertices with \( p \) components. We have noted in EXAMPLE 5A.22 that the sequence \( \{p_n\} \) is a normalized basis sequence for \( g(u) = g_1u + g_2u^2/2! + \cdots \) where \( g_n \) is the number of rooted trees on a set of \( n \) vertices. We did not need to know that \( g_n = n^{n-1} \) to conclude this fact. A little thought reveals that these polynomials must satisfy the identity \( g_n = np_{n-1}(1) \). To see this, one notes that to construct a rooted tree on \( n \) vertices one can select the root (this accounts for the factor \( n \)) and then construct a rooted forest on \( n - 1 \) vertices (this accounts for the factor \( p_{n-1}(1) \)). The forest is just the collection of principal subtrees of the root (CHAPTER 6, DEFINITION 6.28). Direct substitution reveals that the Abel polynomials \( x(x+n)^{n-1} \) also satisfy this relation and hence both have the same power series \( g(u) \). Thus \( e^{xg(u)} \) generates both sequences and, hence, the polynomials are equal. This is an independent and rather curious verification of the fact that \( n^{n-1} \) counts the number of rooted trees on \( n \) vertices!

If \( p_n \) is the normalized basis sequence of \( g(u) \) then the functional equation \( e^{(x+y)g(u)} = e^{xg(u)}e^{yg(u)} \) implies that \( p_n(x+y) = \sum_{k=0}^{n} \binom{n}{k} p_k(x)p_{n-k}(y) \) holds.

Thus the normalized basis sequence of a power series \( g(u) \) is a sequence of polynomials of binomial type in the sense of DEFINITION 5A.33.
5A.33 DEFINITION.

A basis sequence \( \{p_n\} \) (DEFINITION 5A.23) that satisfies \( \sum_{k=0}^{n} \binom{n}{k} p_k(x)p_{n-k}(y) = p_n(x+y) \) for all \( n \) is called a sequence of polynomials of binomial type.

As we previously noted, a normalized basis sequence of a power series \( g(u) = g_1u + g_2u^2/2! + \ldots \) (as usual, \( g_1 \neq 0 \)) is always a sequence of polynomials of binomial type. The converse is also true (EXERCISE 5A.34(5)). For this reason, the theory of exponential generating functions of the form \( e^{xg(u)} \) is sometimes referred to as the "theory of polynomials of binomial type."

5A.34 EXERCISE.

(1) Prove that the number of ordered partitions \( (A_1,A_2,\ldots,A_p) \) of a set of \( n \) elements is given by the multinomial coefficient of NOTATION 5A.14 where \( a_i = |A_i| \) (the \( a_i \) may be zero, in which case \( 0! = 1 \)).

(2) Using LEMMA 5A.24, prove COROLLARY 5A.25.

(3) Let \( G \) denote the set of all power series of the form \( a(u) = a_1u + a_2u^2 + a_3u^3 + \ldots \) where \( a_1 \neq 0 \). Show that \( G \) is a group under composition of power series. (The identity in \( G \) is the trivial power series \( u \).)

(4) Show that \( E - 1, 1 - E^{-1} \), and \( DE^{-a} \) are the basis shift operators for the shift bases \( (x)_n/n! \), \( (x)^{n}/n! \), and \( x(x+an)^{n-1}/n! \) respectively.

(5) Prove that every sequence \( \{p_n\} \) of polynomials of binomial type (DEFINITION 5A.33) is the normalized basis sequence of a power series \( g(u) \) as defined in DEFINITION 5A.27. Hint: First show that \( \{p_n\} \) must in fact be normalized in the sense of DEFINITION 5A.23. Define \( g(u) = g_1u + g_2u^2/2! + \ldots \) by defining \( g_n \) to be the coefficient of \( x \) in \( p_n(x) \).

(6) Each element \( g \in G \) (the group of (3) above) defines a linear transformation \( U_g \) on polynomials. The defining relation is \( U_gx^n = p_n(x) \) where \( e^{xg(u)} = 1 + p_1(x)u + p_2(x)u^2/2! + \ldots \). If \( gh \) denotes the composition of \( g \) and \( h \) (i.e., the group operation of \( G \)) show that \( U_{gh} = U_gU_h \) and thus that the correspondence \( g \rightarrow U_g \) is a "representation" of the group \( G \). The operators \( U_g, g \in G \), are called umbral operators.

(7) The Dirichlet generating function of a sequence \( a_1,a_2,\ldots \) is the formal sum \( \sum_{n=1}^{\infty} a_n n^{-x} \). State the rule for taking products of Dirichlet generating functions analogous to PRODUCT RULES 5A.2 and 5A.11. The Dirichlet generating function for the sequence \( 1,1,\ldots \) is \( \zeta(x) = \sum_{n=1}^{\infty} n^{-x} \) and is called the "Riemann zeta function." A famous unsolved mathematical conjecture called the "Riemann Hypothesis" is concerned with the zeroes of the analytic continuation of \( \zeta(x) \), \( x > 1 \).
5B. THE PRINCIPLE OF INCLUSION-EXCLUSION.

Consider the product \((1 - x_1)(1 - x_2) \ldots (1 - x_n)\) where \(x_1, \ldots, x_n\) are variables. Expanding this product, we get \(\sum_{Q \subseteq \mathcal{P}(\mathbb{N})} (-1)^{|Q|} \prod_{q \in Q} x_q\). The sum is over all subsets \(Q\) of \(\mathbb{N}\). Intuitively, each term in this product arises by choosing a subset \(Q\) of the index set \(1, \ldots, n\) and choosing the factor \(-x_q\) for each \(q \in Q\) and choosing the factor 1 for each \(q \notin Q\). All possible terms are obtained by varying \(Q\) over all possible subsets of \(\mathbb{N}\). Given any set \(S\), we may consider real valued functions \(f_1, f_2, \ldots, f_n\) instead of the variables \(x_1, x_2, \ldots, x_n\). Sums and products of functions are interpreted point-wise: \((f_i + f_j)(t) = f_i(t) + f_j(t)\) and \((f_i f_j)(t) = f_i(t)f_j(t)\). With this interpretation, the above expansion is obviously still true and is called the "principle of inclusion-exclusion," as stated in THEOREM 5B.1.

5B.1 THEOREM (PRINCIPLE OF INCLUSION-EXCLUSION).

Let \(f_1, f_2, \ldots, f_n\) be real valued functions. Then, if \(u\) is the identically 1 function, \(u(t) = 1\) for all \(t\),

\[
\prod_{i=1}^{n} (u - f_i) = \sum_{Q \subseteq \mathcal{P}(\mathbb{N})} (-1)^{|Q|} \prod_{q \in Q} f_q.
\]

The most elementary combinatorial applications of THEOREM 5B.1 occur when the functions \(f_i\) are chosen to be characteristic functions \(f_{P_i}\) of subsets \(P_1, \ldots, P_n\) of a set \(P\) (see EXERCISE 1.13(6)). In this case \(u - f_{P_i}\) is the characteristic function of the set \(P - P_i\) (the domain of all functions is \(P\)). Thus, the product on the left-hand side of the identity of THEOREM 5B.1 is the characteristic function of the set of all elements of \(P\) that do not belong to any of the \(P_i, i = 1, \ldots, n\). It is an interesting fact that in some cases the right-hand side of THEOREM 5B.1 is easier to compute than the left-hand side. Note that if \(f_A\) is a characteristic function of a subset \(A \subseteq P\), then \(\sum_{t \in P} f_A(t) = |A|\). Thus, by evaluating both sides of the identity of THEOREM 5B.1 at \(t\) and summing over all \(t\), we obtain, in the case \(f_i = f_{P_i}\), COROLLARY 5B.2.

5B.2 COROLLARY (ENUMERATIVE INCLUSION-EXCLUSION).

\[
\left| \bigcap_{i=1}^{n} (P - P_i) \right| = \sum_{Q \subseteq \mathcal{P}(\mathbb{N})} (-1)^{|Q|} \left| \bigcap_{q \in Q} P_q \right|.
\]

A standard example of the application of COROLLARY 5B.2 is to the problem of derangements (also considered in EXAMPLE 5A.12).
5B.3 EXAMPLE: DERANGEMENTS AND INCLUSION-EXCLUSION.

We let $P$ denote the set $S_n$ of all permutations of $n$. Let $D(n)$ be the derangements as defined in EXAMPLE 5A.12. Let $P_i$ be the set of all permutations $\sigma$ in $S_n$ such that $\sigma(i) = i$. $P_i$ is defined for $i = 1, \ldots, n$. Clearly, $|\bigcap_{q \in Q} P_q| = (n - |Q|)!$. This example is typical in that this quantity does not depend on the set $Q$ but only on $|Q|$. Since $\bigcap_{i=1}^{n} (P - P_i) = D(n)$, we have from COROLLARY 5B.2 the identity $|D(n)| = \sum_{Q \subseteq \mathbb{N}} (-1)^{|Q|} (n - |Q|)! = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)! = n! \sum_{k=0}^{n} (-1)^k/k!$. Again, we note that this is approximately $n!e^{-1}$. This approach should be compared carefully with that of EXAMPLE 5A.12.

For some applications it is convenient to have a slight generalization of THEOREM 5B.1.

5B.4 COROLLARY.

Let $f_1, \ldots, f_n$ be as in THEOREM 5B.1. Let $T \subseteq \mathbb{n}$ and let $T^c = n - T$. Then

$$\prod_{i \in T} f_i \prod_{i \in T^c} (u - f_i) = \sum_{R \supseteq T} (-1)^{|R| - |T|} \prod_{i \in R} f_i.$$ 

**Proof.** The left side equals (using THEOREM 5B.1)

$$\prod_{i \in T} f_i \sum_{Q \subseteq T^c} (-1)^{|Q|} \prod_{q \in Q} f_q.$$ 

If we set $R = T \cup Q$, this expression becomes $\sum_{R \supseteq T} (-1)^{|R| - |T|} \prod_{i \in R} f_i$, which is the right side of the identity.

As above, we may let $f_i = f_{P_i}$ in COROLLARY 5B.4. In this case, the left-hand side of COROLLARY 5B.4 becomes the cardinality of the set (which we call "IN(T)") of all elements of $P$ that belong to $P_i$ if and only if $i \in T$. This leads to DEFINITION 5B.5.

5B.5 DEFINITION.

Let $P_i, i = 1, \ldots, n$, be subsets of $P$. Let $T \subseteq \mathbb{n}$ be a subset of $n$. Define $\text{IN}(T) = \{p: p \in P, p \in P_i \text{ if and only if } i \in T\}$ and define $\text{IN}(k) = \bigcup_{|T| = k} \text{IN}(T)$ where the union is over all subsets $T$ of $n$ with $|T| = k$. $\text{IN}(k)$ is the set of all elements of $P$ that belong to exactly $k$ of the sets $P_i$. 

Digested by Google
5B.6 COROLLARY.

Let $P_i$, $i = 1, \ldots, n$, be subsets of $P$ and let $T \subseteq n$. Let $IN(T)$ and $IN(k)$ be as in DEFINITION 5B.5. Then the following identities are valid

1. $|IN(T)| = \sum_{R \supseteq T} (-1)^{|R|-|T|} \sum_{i \in R} |P_i|.$

2. $|IN(k)| = \sum_{r=k}^{n} \binom{r}{k} (-1)^{r-k} \sum_{|R|=r} \sum_{i \in R} |\cap_{i \in R} P_i|$ where the sum is over all subsets $R \subseteq n$ with $|R| = r$.

Proof. Identity (1) follows from COROLLARY 5B.4 just as COROLLARY 5B.2 follows from THEOREM 5B.1. Identity (2) requires a change of order of summation as follows:

$$|IN(k)| = \sum_{|T|=k} \sum_{R \supseteq T} (-1)^{|R|-|T|} \sum_{i \in R} |P_i|$$
$$= \sum_{|R|\geq k} \binom{|R|}{k} (-1)^{|R|-k} \sum_{i \in R} |P_i|$$
$$= \sum_{r=k}^{n} \binom{r}{k} (-1)^{r-k} \sum_{|R|=r} \sum_{i \in R} |P_i|$$

which is the right-hand side of identity (2).

5B.7 EXAMPLE OF COROLLARY 5B.6 APPLIED TO DERANGEMENTS.

As in EXAMPLE 5B.3, let $P = S_n$ and let $P_i = \{\sigma: \sigma(i) = i\}$. The set $IN(k)$ is the set of all permutations that fix exactly $k$ elements of $n$. We recall from EXAMPLE 5B.3 that $|\cup_{i \in R} P_i| = (n-|R|)!$ and hence from COROLLARY 5B.6 we have

$$IN(K) = \sum_{r=k}^{n} (-1)^{r-k} \binom{r}{k} \sum_{|R|=r} (n-|R|)!$$
$$= \sum_{r=k}^{n} (-1)^{r-k} \binom{r}{k} \binom{n}{r} (n-r)!$$
$$= \frac{n!}{k!} \sum_{r=k}^{n} \frac{(-1)^{r-k}}{(r-k)!}$$
$$= \frac{n!}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \approx \frac{n!}{k!} e^{-1} (n-k \text{ large enough}).$$
5B.8 EXAMPLE OF COROLLARY 5B.6 APPLIED TO DIVISIBILITY CONDITIONS.

Let \( P = m \) and let \( \{a_1, \ldots, a_n\} \) be a set of pairwise relatively prime positive integers (two integers are \textit{relatively prime} if their greatest common divisor is 1). Let \( P_i = \{t: t \in m \text{ and } a_i \text{ divides } t\} \). We also write \( a_i|t \) for \text{“}a_i \text{ divides } t\text{.”} For any subset \( Q \subseteq n, \cap P_i = \{t: t \in m \text{ and } \prod_{i \in Q} a_i|t\} = \prod_{i \in Q} a_i \), \( \prod_{i \in Q} a_{i_1} \ldots \prod_{i \in Q} a_{i_k} \) where \( k = |x| \) (see DEFINITION 2.2) with \( x = \prod_{i \in Q} a_i \). For this example \( \text{IN}(k) \) is the set of all integers of \( m \) divisible by exactly \( k \) of the integers in the set \( \{a_1, \ldots, a_n\} \). Applying COROLLARY 5B.6(2), we see that

\[
|\text{IN}(k)| = \sum_{r=k}^{n} \binom{r}{k} (-1)^{r-k} \sum_{|R|=r} \left( \frac{m}{\prod_{i \in R} a_i} \right).
\]

\( R \) is a subset of \( n \). If each of the integers \( a_i \) is a divisor of \( m \) then \( \prod_{i \in R} a_i \) is an integer for all \( R \). In this case, \( |\text{IN}(0)| = m \prod_{t=1}^{n} \left( 1 - \frac{1}{a_t} \right) \). The number \( |\text{IN}(0)| \) is sometimes denoted by \( \varphi(m) \) and is called the \textit{Euler \varphi-function}.

5B.9 EXERCISE.

(a) How many ways are there of placing eight nonattacking rooks on an \( 8 \times 8 \) chessboard with exactly one rook on the white diagonal? \textit{Hint:} Rook placements correspond in a natural way to permutations (elements of \( S_8 \)). A rook on the white diagonal corresponds to a fixed point of such a permutation.

(2) How many integers between one and 1000 are divisible by exactly \( k \) of the integers 3, 5, and 7 for \( k = 0, 1, 2, \) and 3?

We now develop a more intricate example of the principle of inclusion-exclusion. The problem we consider is called the \textit{“problème des ménages.”} Suppose \( n \) couples are to be seated at a circular table. The \( n \) husbands take alternate seats. At this point, how many ways are there to seat the \( n \) wives such that no wife sits next to her own husband? A seating may be described by a permutation \( \varphi \in S_n \) by specifying that wife \( \varphi(i) \) is to be seated to the left of gentleman \( i \). We must have \( \varphi(i-1) \neq i \) and \( \varphi(i) \neq i \) for all \( i \) (arithmetic mod \( n \)). We thus may consider sets \( P_{2i-1} = \{\varphi: \varphi(i-1) = i\} \) and \( P_{2i} = \{\varphi: \varphi(i) = i\}, i = 1, \ldots, n \). Our problem is to determine the number \( |S_n - \bigcup_{j=1}^{2n} P_j| \). The solution is described in EXERCISE 5B.10.
5B.10 EXERCISE.

(See the previous paragraph for statement of problem.)

1. Show that for $Q \subseteq 2n$, $\sum_{i \in Q} |P_i| = 0$ if $Q$ contains two consecutive integers mod $2n$ and $\sum_{i \in Q} |P_i| = (n - |Q|)!$ if $Q$ does not contain two consecutive integers mod $2n$ (assume $n > 1$).

2. Show that for $k \leq n$ the number of ways of selecting a set $Q$ with $|Q| = k$ from $2n$ such that no two elements of $Q$ are consecutive mod $2n$ is

$$N^*(2n,k) = \frac{2n}{2n-k} \binom{2n-k}{k}.$$

Hint: Let $N(m,k)$ denote the number of sequences of zeroes and ones with length $m$, $k$ ones, and no two ones consecutive. Show that $N(m,k) = \binom{m-k+1}{k}$. For $m = 4$, $k = 2$, $1001$ is such a sequence. Let $N^*(m,k)$ denote the number of such sequences where "consecutive" now applies to the position 1 and $m$ (consecutive mod $m$). Thus $1001$ would not be included in this count. Show $N^*(m,k) = N(m-1,k) + N(m-3,k-1) = \frac{m}{m-k} \binom{m-k}{k}$. The correspondence between these sequences and the set $Q$ is obtained by considering the characteristic function $f_Q$.

3. Show that the number of solutions to the "problème des ménages" is given by

$$\sum_{k=0}^{n} (-1)^k(n-k)! \frac{2n}{2n-k} \binom{2n-k}{k}, \ n > 1.$$

Certain applications of the principle of inclusion-exclusion require a slight generalization of COROLLARIES 5B.2 and 5B.6. Both of these results were obtained from THEOREM 5B.1 by choosing the functions $f_1, \ldots, f_n$ to be characteristic functions $f_{P_1}, \ldots, f_{P_n}$. We then used the fact that, for any subset $A \subseteq P$, $\sum_{t \in P} f_A(t) = |A|$ to pass from THEOREM 5B.1 to COROLLARIES 5B.2 and 5B.6. If $w: P \to X$ is any function from $P$ to an algebraic structure $X$ where addition and subtraction are defined (an abelian group, ring, etc.) we can define $w(A) = \sum_{t \in A} w(t) = \sum_{t \in P} w(t)f_A(t)$. In the case where $w(t) = 1$ for all $t \in P$ we get $w(A) = |A|$. We call $w$ a "weight function." COROLLARIES 5B.11 and 5B.12 are obvious extensions of COROLLARIES 5B.2 and 5B.6 to the case of weight functions.
5B.11 COROLLARY (WEIGHTED INCLUSION-EXCLUSION).

Using the notation of COROLLARY 5B.2 we have

\[ w\left( \bigcap_{i=1}^{n} (P - P_i) \right) = \sum_{\mathcal{Q} \subseteq \mathcal{P}} (-1)^{|\mathcal{Q}|} w\left( \bigcap_{q \in \mathcal{Q}} P_q \right). \]

5B.12 COROLLARY.

Using the notation of COROLLARY 5B.6 we have

(1) \[ w(\text{IN}(T)) = \sum_{R \supseteq T} (-1)^{|R| - |T|} \sum_{i \in R} w(\bigcap_{i} P_i). \]

(2) \[ w(\text{IN}(k)) = \sum_{r=k}^{n} \binom{r}{k} (-1)^{r-k} \sum_{|R| = r} \sum_{i \in R} w(\bigcap_{i} P_i). \]

We now give an example of the use of the weighted principle of inclusion-exclusion.

5B.13 EXAMPLE OF WEIGHTED INCLUSION-EXCLUSION AND PERMANENTS.

In this example, we let \( P = \mathcal{P} \) be the set of all functions from \( m \) to \( n \) where we assume that \( m \leq n \). Let \( P_i, i = 1, \ldots, n \), be defined by \( P_i = \{ f : i \in \text{image}(f) \} \).

For \( k = n - m \) the set \( \text{IN}(k) \) is the set of injective mappings from \( m \) to \( n \) (note that any \( f \in \text{IN}(k) \) has \( |\text{image}(f)| = n - k = m \) and hence must be an injection). Let \( A = (a_{ij}) \) be an \( m \times n \) matrix. For each \( f \) in \( P \) we define the weight of \( f \) by

\[ w(f) = \prod_{t=1}^{n} a_{i_t f(t)}. \]

We define the permanent of \( A \), \( \text{per}(A) \), to be \( w(\text{IN}(k)) \) where \( k = n - m \). In other words, \( \text{per}(A) = \sum_{f \in \text{IN}(k)} \prod_{t=1}^{n} a_{i_t f(t)} \) where the sum is over all injections from \( m \) to \( n \). If \( m = n \) the permanent is the same as the determinant except that the determinant uses the weight function \( w(f) = \text{sgn}(f) \prod_{t=1}^{n} a_{i_t f(t)} \)

where the \( \text{sgn}(f) \) is \(+1\) for even permutations and \(-1\) for odd permutations. The properties of the permanent are such that it is generally much more difficult to compute than the determinant. In some cases, COROLLARY 5B.12(2) provides the best means for computing the permanent as we now explain. We have immediately, from COROLLARY 5B.12(2), that \( \text{per}(A) = \sum_{r=k}^{n} \binom{r}{k} (-1)^{r-k} W_A(r) \)

where \( W_A(r) = \sum_{|R| = r} \sum_{i \in R} w(\bigcap_{i} P_i) \) with \( w \) defined as above. But, the set \( \bigcap_{i} P_i \) is just the set of all functions \( f \) with \( \text{image}(f) \) contained in \( n - R \). In other words, the set of all functions from \( m \) to \( n - R \), \( (n - R)^m \). Thus, it is easy to see that
\[ w(\cap P_i) = \prod_{i \in R} \sum_{t=1}^{m} a_{ij}. \] This formula can be used in certain instances to compute \( W_A(r) \). As an example, consider the \( 3 \times 4 \) matrix

\[
A = \begin{pmatrix}
1 & -1 & -1 & 1 \\
1 & 1 & 1 & -1 \\
1 & -1 & 1 & -1
\end{pmatrix}.
\]

There are four subsets \( R \subseteq 4 \) of size 1: \{1\}, \{2\}, \{3\}, \{4\}. For \( r = 1 \), we have (using the above formula)

\[
W_A(1) = (-1)(+1)(-1) + (+1)(+1)(+1) + (+1)(+1)(-1) + (-1)(+3)(+1) = -2.
\]

Similarly, we find that \( W_A(2) = 0 \), \( W_A(3) = 2 \) and hence that \( \text{per}(A) = W_A(1) - \binom{2}{1} W_A(2) + \binom{3}{1} W_A(3) = -2 + 6 = 4 \). This method should be compared with the direct method of computing \( \text{per}(A) \) (using the definition).

5C. MÖBIUS INVERSION.

We now present a more general classical technique than the method of inclusion-exclusion. The expository article of Bender and Goldman (1975) referred to at the end of this chapter contains many examples of this technique. We give a brief presentation of the most important concepts, and refer the reader to this article and the other references listed in the section on MÖBIUS INVERSION for further study. The reader should also review the material at the beginning of Chapter 1 through EXERCISE 1.13.

We refer to an ordered set \( P = (S, \leq) \) as a "poset." We refer to elements \( x, y, \ldots \) of the set \( S \) of the poset \( P \) as "elements of the poset \( P \)" and write \( x \in P, y \in P, \ldots \).

5C.1 DEFINITION.

Let \( P \) be a poset, \( x, y \in P, x \leq y \). Define the interval \([x, y]\) to be the set \( \{z : x \leq z \leq y\} \).

5C.2 EXAMPLE OF AN INTERVAL IN A POSET.

Consider the poset of all subsets of \( A, \mathcal{P}(A) \), where \( A = 4 \) (EXAMPLE 1.9(3)). The interval \([\{1,2\}, \{1,2,3,4\}] = \{\{1,2\}, \{1,2,3\}, \{1,2,4\}, \{1,2,3,4\}\} \). An interval \([x, y]\) in a poset \( P = (S, \leq) \) is itself a poset \( ([x, y], \leq) \) with the same order relation as \( P \) restricted to \([x, y] \times [x, y]\). We refer to this poset as the "subposet \([x, y]\)." The Hasse diagram for the subposet of this example is shown in FIGURE 5C.3.
5C.3 HASSE DIAGRAM OF AN INTERVAL.

![Hasse Diagram](image)

A poset $P$ is "locally finite" if every interval has finitely many elements.

5C.4 DEFINITION.

A poset $P$ is *locally finite* if $|[x,y]|$ is finite for all $x,y \in P$.

Henceforth, "poset" will mean "locally finite poset."

Let $D$ be any set and let $R$ be the real numbers. Then the set of all functions from $D$ to $R$, which we denote by $R^D$, is a vector space over $R$ under the usual operations of function addition and scalar multiplication. A vector space together with a rule for multiplying vectors is called an *algebra*. Formally, the rule of multiplication must satisfy certain axioms, but these axioms will be trivially true for the examples we shall be considering. The reader already knows a number of such algebras (the vector space $R^D$ under pointwise multiplication was considered in connection with the INCLUSION-EXCLUSION section of this chapter, the vector space of polynomials under polynomial multiplication, the vector space of $n \times n$ matrices under matrix multiplication). We now take the set $D$ to be the set $i(P)$ of all intervals of a (locally finite) poset $P$. The vector space $R^{i(P)}$ will be called the "incidence vector space" of the poset $P$. We now define a multiplication on $R^{i(P)}$.

5C.5 DEFINITION.

Let $P$ be a poset, and let $f$ and $g$ be elements of $R^{i(P)}$. The *convolution* of $f$ and $g$, denoted by $f * g$ is defined by

$$f * g([x,y]) = \sum_{x \leq z \leq y} f([x,z])g([z,y]) .$$

The vector space $R^{i(P)}$ with the convolution rule of multiplication is called the *incidence algebra* of the poset $P$. 

Digitized by Google
Incidence algebras of finite posets are all isomorphic to subalgebras of \( n \times n \) matrices where \( n = |P| \). In fact the matrices may always be taken to be upper triangular. We discuss this aspect of incidence algebras shortly. First, however, we shall give direct elementary derivations of some of the properties of the incidence algebra of a poset. This is perhaps the best way to get a feeling for how the structure of the poset relates to the structure of the algebra.

5C.6 DEFINITION.

The function \( \delta \in \mathcal{R}(P) \) defined by \( \delta([x,x]) = 1 \) for all \( x \in P \) and \( \delta([x,y]) = 0 \) for all \( x \neq y, x,y \in P \), is a multiplicative identity \( (\delta*f = f*\delta = f \text{ for all } f \in \mathcal{R}(P)) \) for the incidence algebra. The function \( \delta \) is called the delta function.

5C.7 LEMMA.

The incidence algebra is associative: \( f*(g*h) = (f*g) * h \).

Proof. Verify that \( (f*g)*h = f*(g*h) \). We have

\[
((f*g)*h)[x,y] = \sum_{x \leq z \leq y} (f*g)[x,z]h[z,y] = \sum_{x \leq z \leq y} \sum_{x \leq w \leq z} f[x,w]g[w,z]h[z,y].
\]

We interchange the order of summation to obtain

\[
\sum_{x \leq w \leq y} \sum_{w \leq z \leq y} f[x,w]g[w,z]h[z,y] = \sum_{x \leq w \leq y} f[x,w](\sum_{w \leq z \leq y} g[w,z]h[z,y]) = \sum_{x \leq w \leq y} f[x,w](g*h)[w,y] = f*(g*h)[x,y].
\]

An element \( f \) in \( \mathcal{R}(P) \) is a “unit” or “invertible element” if there exists an element \( g \) such that \( f*g = g*f = \delta \). LEMMA 5C.8 characterizes these elements.

5C.8 LEMMA.

An element \( f \in \mathcal{R}(P) \) is invertible if and only if \( f([x,x]) \neq 0 \) for all \( x \in P \).

Proof. Assume \( f \) is invertible. For two functions \( f \) and \( g \) in \( \mathcal{R}(P) \) we have \( f*g[x,x] = f[x,x]g[x,x] \). Thus if \( f*g = \delta \), then \( f[x,x] \neq 0 \).

Assume \( f[x,x] \neq 0 \). We define \( g \) inductively on the cardinality of the interval.

If \( x \in P \) define \( g[x,x] = \frac{1}{f[x,x]} \). Assume we have defined \( g \) on all intervals \([x,y], ||[x,y]| = n \). Let \( ||[x,y]| = n + 1 \). For \( f*g[x,y] = \delta[x,y] = 0 \) we must have

\[
f[x,x]g[x,y] + \sum_{x < z \leq y} f[x,z]g[z,y] = 0.
\]
Thus we define (note that \(|x,z|\) and \(|z,y|\) are \(\leq n\))

\[
g(x,y) = \frac{-1}{f(x,x)} \sum_{x \leq z \leq y} f(x,z)g(z,y).
\]

With \(g(x,y)\) so defined we have constructed a right inverse for \(f\) (i.e., \(f \ast g = \delta\)). Thus any element \(f\) with \(f(x,x) \neq 0\) has a right inverse \(g\), \(f \ast g = \delta\). But then \((g \ast f) \ast g = g \ast (f \ast g) = g\). Since \(f(x,x)g(x,x) = \delta(x,x) = 1\), \(g(x,x) \neq 0\) so let \(h\) be the right inverse of \(g\). Thus \(g \ast f = (g \ast f) \ast (g \ast h) = ((g \ast f) \ast g) \ast h = g \ast h\) or \(g \ast f = \delta\). Thus \(g\) is also a left inverse of \(f\) and hence \(g\) is the inverse of \(f\).

**Lemma 5C.8** provides a constructive method for defining the inverse of any invertible \(f\), although in most instances a more concise representation of the inverse is required. This method is essentially "Gaussian elimination" of matrix theory.

Combinatorially, the most important functions in the incidence algebra are the zeta function and its inverse, the Möbius function. These functions are defined in **Definition 5C.9**.

**5C.9 Definition.**

The function \(\zeta \in R^{(P)}\) which is identically 1 for all intervals \([x,y] \in i(P)\) is called the zeta function. The inverse \(\mu\) of the zeta function is called the Möbius function of \(P\).

As \(\zeta([x,y]) = 1\) for all \([x,y] \in i(P)\), \(\zeta\) is clearly invertible by **Lemma 5C.8**. We shall be interested in obtaining simple formulas for the value of \(\mu([x,y])\) in terms of \(x\) and \(y\) for interesting posets \(P\). For example, we shall see that for the poset \(\mathcal{P}(A)\) of all subsets of a set \(A\), \(\mu([x,y]) = (-1)^{y-x}\) where \(|y-x|\) is the number of elements in the subset \(y\) but not in the subset \(x\). Before going into this topic in detail, we make some simple observations about the Möbius function \(\mu\). Note that if \(x \in P\), then \(\mu(x,x) = \frac{1}{\zeta(x,x)} = 1\). If \(|[x,y]| = 2\) then \(\mu \ast \zeta([x,y]) = 0 = \delta([x,y])\) implies that \(\mu(x,x) + \mu(x,y) = 0\) so \(\mu(x,y) = -1\). If \(|[x,y]| = 3\) then again, if \(x < z < y\), then \(\mu(x,x) + \mu(x,z) + \mu(x,y) = 0\) so \(\mu(x,y) = 0\). If \(|[x,y]| = 4\) there are two possibilities shown in **Figure 5C.10**. In **Figure 5C.10(a)** \(\mu([x,y]) = 0\) but in **Figure 5C.10(b)** \(\mu([x,x]) + \mu([x,z]) + \mu([x,z_2]) = 1 + (-1) + (-1) = -1\) so \(\mu([x,y]) = +1\).

Thus, one sees that as the situation grows more complex the nature of \(\mu\) depends very much on the structure of the poset and more sophisticated techniques will be required to characterize \(\mu\). One should note that if \(P\) is linearly ordered \((x,y \in P\) implies that \(x \leq y\) or \(y \leq x\)) then the function \(f\) defined by \(f([x,x]) = 1\) for all \(x\), \(f([x,y]) = -1\) if \(|[x,y]| = 2\), and \(f([x,y]) = 0\) if \(|[x,y]| > 2\) inverts \(\zeta\) and hence (by uniqueness of inverses) must be equal to \(\mu\).
5C.10 TWO INTERVALS OF CARDINALITY FOUR.

![Diagram of two intervals with points labeled x, y, z₁, z₂](image)

Figure 5C.10

We have already mentioned that, as an algebra, the incidence algebra of a finite poset is isomorphic to a subalgebra of $n \times n$ upper triangular matrices. We now indicate why this is so. Consider the $6 \times 6$ matrix of FIGURE 5C.11. The rows and columns of this matrix are indexed by the elements of the poset $\mathcal{P}(3)$ of all subsets of $3$. In order to write down such a matrix, the elements of the poset must be linearly ordered. We have chosen the linear order on $\mathcal{P}(3)$ such that, if the subset $x$ is contained in the subset $y$, then $x$ comes before $y$ in the linear order. This can be done in general, and is not difficult to show for the case of finite posets (EXERCISE 5C.12(1)). When such a linear order is specified, the matrix $Z$, called the incidence matrix, corresponding to that of FIGURE 5C.11 is always upper triangular. Note that the matrix $Z$ of FIGURE 5C.11 has a 1 in position $(x,y)$ if and only if $x \subseteq y$.

Thus the incidence matrix is just one possible way of representing the zeta function of the poset. Let $\mathcal{A}(P)$ denote the set of all matrices $M$ with $M(x,y) = 0$ whenever $Z(x,y) = 0$. Note that the convolution product of DEFINITION 5C.5 when applied to these matrices as members of $R^{\mathcal{A}(P)}$ (which they are, because $M$ is the function $[x,y] \rightarrow M(x,y)$) is just the standard rule for multiplying matrices. The product of two matrices in $\mathcal{A}(P)$ is again a matrix in $\mathcal{A}(P)$. It is easily seen that the matrix algebra $\mathcal{A}(P)$ is isomorphic to the incidence algebra $R^{\mathcal{A}(P)}$. We skip the algebraic formalities, as the result should have strong intuitive appeal at this point. With the matrix algebra model in mind, some of our previous results become obvious to the reader familiar with a little matrix theory. For example, LEMMA 5C.7 is just the statement that matrix multiplication is associative. LEMMA 5C.8 states the obvious fact that an upper triangular matrix is invertible if and only if all of its diagonal entries are nonzero. It turns out to
be useful to have both the matrix point of view and the more local point of view in mind when studying the incidence algebras of posets. The standard linear algebraic ideas are not enough to answer the questions that arise in the study of incidence algebras. The special structures of the various posets of combinatorics must be exploited. We shall now study some specific examples where we try to give concise formulas for the entries of the inverse of $Z$. Using Cramer's rule or Gaussian elimination or other standard linear algebraic results is not enough to provide these results.

**5C.11 AN INCIDENCE MATRIX.**

$$
\begin{array}{cccccccc}
   & \phi & 1 & 2 & 3 & 12 & 13 & 23 & 123 \\
\phi & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\
2 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\
3 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\
12 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
13 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
23 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
123 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
$$

Figure 5C.11

**5C.12 EXERCISE.**

1. Let $S$ be a finite set with order relation $\leq$. Give an algorithm that lists the elements of $S$ sequentially: $x_1, x_2, \ldots, x_n$, such that $x_q < x_p$ implies $x_p$ is to the right of $x_q$ in the list ($q < p$ as integers). Call such a listing order compatible for the given order relation $\leq$. The column headings (read left to right) for the matrix $Z$ of FIGURE 5C.11 specify an order compatible listing of the elements of the poset $\mathcal{P}(3)$. An order compatible listing of $S$, when used to define the incidence matrix $Z$ of the poset $P = (S, \leq)$, forces $Z$ to the upper triangular, as in FIGURE 5C.11.

2. Find the inverse of the matrix $Z$ of FIGURE 5C.11 using any one of the standard methods of elementary linear algebra.

We now show how one computes the Möbius function of a product of posets from those of its factors. This result is the most important elementary tool for computing formulas for Möbius functions of posets. Let $\{P_i; i \in I\}$ be a family
of posets indexed by the possibly infinite set I. Suppose $P_i = (S_i, \leq_i)$. Consider the Cartesian product $\times_{i \in I} S_i = S$. If $f \in S$ then $f(i)$ denotes the "component" of $f$ in $S_i$. For $f$ and $g$ in $S$, we define $f \leq g$ if $f(i) \leq_i g(i)$ for all $i \in I$. $P = (S, \leq)$ is easily seen to be a poset (EXERCISE 1.13(3)). We write $P = \times_{i \in I} P_i$ and call $P$ the product of the posets $P_i$, $i \in I$. If the index set $I$ is infinite then $P$ may not be locally finite even if all of the $P_i$ are locally finite (almost any example shows this). For this reason, we fix some element $f_0 \in S$ and let $S_0 = \{ f : [i: f(i) \neq f_0(i)] < \infty \}$. $S_0$ is the set of all elements of $S$ that are equal to $f_0$ except for finitely many components. Define $P_0 = (S_0, \leq)$. The poset $P_0$ is clearly locally finite if all of the $P_i$ are locally finite. For example, let $P_i = \{0, 1, 2, \ldots\} = N$ be the non-negative integers ordered in the usual manner. Let $I = \{1, 2, 3, \ldots\}$ be the positive integers. Then $P = \times_{i \in I} P_i = N^I$ corresponds to all functions from $I$ to $N$. If $f_0$ is chosen to be the identically zero function, the $P_0$ consists of all functions that are zero except at finitely many values of $I$. The general result we need is stated in THEOREM 5C.13.

5C.13 THEOREM.

Let $f \leq g$ be in $P_0$ (defined in the previous paragraph) and let $\mu$ denote the Möbius function of $P_0$. Then $\mu((f, g)) = \prod_{i \in I} \mu_i([f(i), g(i)])$ where $\mu_i$ is the Möbius function of $P_i$.

Proof. Let $[f, g] \in i(P_0)$ be an interval of $P_0$. Define $\tau[f, g] = \prod_{i \in I} \mu_i([f(i), g(i)])$. This product is finite. Consider $\tau \ast \zeta[f, g] = \sum_{f = h \leq g} \tau[f, h] = \sum_{f = h \leq g} \prod_{i \in I} \mu_i([f(i), h(i)])$.

Let $\mu_i([f(i), j]) = a_{ij}$ and let $f(i) = b_i$ and $g(i) = c_i$. Then $\tau \ast \zeta[f, g] = \sum_h \prod_{i \in I} a_{ih(i)}$ where the sum is over all $h$ in $\times_{i \in I} [b_i, c_i]$. By the standard "sum-product interchange" rule, $\tau \ast \zeta[f, g] = \prod_{i \in I} (\sum_{b_i \leq j \leq c_i} a_{ij})$. Thus $\tau \ast \zeta[f, g] = \prod_{i \in I} (\sum_{f(i) \leq j \leq g(i)} \mu_i([f(i), j])) = \prod_{i \in I} \mu_i \ast \zeta_i[f(i), g(i)] = \prod_{i \in I} \delta_i[f(i), g(i)] = \delta[f, g]$ where $\zeta_i$ and $\delta_i$ are the zeta and delta functions of $P_i$ and $\delta$ is the delta function of $P_0$. Thus $\tau$ is the inverse of $\zeta$ and hence, by uniqueness of the inverse, $\tau = \mu$.

We now consider three important examples of the use of THEOREM 5C.13.

5C.14 THE MÖBIUS FUNCTION OF THE POSET OF SUBSETS.

Consider the set of all subsets $\mathcal{P}(D)$ of a finite set $D$ with set inclusion as the order relation. The poset $\mathcal{P}(D)$ is obviously order isomorphic (DEFINITION 1.54) to the set of functions $\{0, 1\}^D$ where each $Z \in \mathcal{P}(D)$ is associated with its
characteristic function $f_Z$. We apply THEOREM 5C.13 with $P_i = \{0,1\}$ (with $0 < 1$) for all $i \in D = I$. $P_i$ is linearly ordered so its Möbius function is given by (see the paragraph following DEFINITION 5C.9) $\mu_i(x,y) = 1$, $\mu_i(x,y) = -1$ if $x < y$. Since $D$ is finite, we take $P = P_0 = \{0,1\}^D$. The identity of THEOREM 5C.13 becomes $\mu[A,B] = \mu[f_A,f_B] = \prod_{i \in A} \mu_i[f_A(i),f_B(i)] = (-1)^{|B-A|}$. Thus $\mu[A,B] = (-1)^{|B-A|}$ is the Möbius function of the poset of subsets.

5C.15 THE MÖBIUS FUNCTION OF THE INTEGERS UNDER DIVISIBILITY.

Let $I = \{p_1, p_2, p_3, \ldots\}$ be the prime numbers in their natural order as integers. Let $N$ be the non-negative integers, and consider the poset $P = N^I$. Using the function $f_0(p_i) = 0$ for all $p_i \in I$, define $P_0$ as above in connection with THEOREM 5C.13. The poset $P_0$ is easily seen to be order isomorphic to the integers under divisibility (using the unique factorization of integers as powers of primes). Thus each $f \in P_0$ corresponds to the integer $n_f = \prod_{i=1}^{\infty} p_i^{f(p_i)}$ where this product has finitely many terms not equal to 1. Using THEOREM 5C.13, we have that $\mu[f,g] = \prod_{i \in I} \mu_i[f(p_i),g(p_i)]$. Since $N$ is linearly ordered, its Möbius function is as described in the paragraph following DEFINITION 5C.9. Hence, $\mu[f,g] = 0$ if $g(p_i) - f(p_i) \geq 2$ for any prime $p_i$. Otherwise, $\mu[f,g] = (-1)^t$ where $t$ is the number of primes $p_i$ for which $g(p_i) - f(p_i) = 1$. Another way to describe this Möbius function is to form the quotient $n_g/n_f$ of the integers $n_g$ and $n_f$ associated with $g$ and $f$. If this number is divisible by the square of a prime, then $\mu[f,g]$ is 0. If this number is "square free" then $\mu[f,g] = (-1)^t$ where $t$ is the number of primes (all to the power 1) in its prime factorization. This is the classical description of the Möbius function of number theory.

Our final application of THEOREM 5C.13 will be to the poset of partitions of a set. Before doing this we discuss briefly the most important combinatorial use of the Möbius function, namely, the inversion of finite series. Suppose that $g$ and $f$ are in the incidence algebra of a finite poset $P$. Suppose also that $g = f \ast \zeta$ where $\zeta$ is the zeta function of the incidence algebra. Expressed as a summation, this identity becomes $g(x,y) = \sum_{x \leq z \leq y} f(x,z)$. The identity $g = f \ast \zeta$ implies that $f = g \ast \mu$ where $\mu$ is the Möbius function of the poset. In terms of sums this becomes $f(x,y) = \sum_{x \leq z \leq y} g(x,z) \mu(z,y)$. A similar rule holds if $g = \zeta \ast f$. We thus have LEMMA 5C.16.

5C.16 LEMMA.

Let $f$ and $g$ be functions from the intervals $i(P)$ of a poset $P$ to $R$. The following rules for inverting finite sums apply:
(1) If \( g(x,y) = \sum_{x \leq z \leq y} f(x,z) \) then \( f(x,y) = \sum_{x \leq z \leq y} g(x,z) \mu[z,y] \).

(2) If \( g(x,y) = \sum_{x \leq z \leq y} f(z,y) \) then \( f(x,y) = \sum_{x \leq z \leq y} \mu[x,z] g(z,y) \).

**5C.17 EXAMPLE OF INVERSION OF FINITE SUMS.**

Let \( P \) be the positive integers with the usual linear order. Suppose that we are given a function \( g \) in the incidence algebra of \( P \), \( g(x,y) = xy^2 \). We define a function \( f \) on \( \{x,y\} : x \leq y \} \) by the finite series \( xy^2 = \sum_{x \leq z \leq y} f(x,z) \). Clearly, \( f[x,x] = g[x,x] = x^3 \). Using LEMMA 5C.16(1) we see that for \( x < y \), \( f[x,y] = \sum_{x \leq z \leq y} xz^2 \mu[z,y] = x(y-1)^2 \mu[y-1,y] + xy^2 \mu[y,y] = -x(y-1)^2 + xy^2 = x(2y-1) \). Thus, given the explicit formula \( g(x,y) = xy^2 \) for \( g \) together with the implicit definition of \( f \) as a finite sum, we obtain, using the Möbius function of the poset, an explicit formula for \( f \). This is a typical application of "Möbius inversion" of a finite sum. The Möbius function in this example is trivial, but we shall now consider some more complex examples.

There is a variation on LEMMA 5C.16 that is sometimes confusing to the beginner in this subject. The functions \( f \) and \( g \) of LEMMA 5C.16 are given directly as functions from the intervals \( i(P) \) of the poset \( P \) to \( R \). In many applications of LEMMA 5C.16 the functions \( f \) and \( g \) are functions from \( P \) to \( R \) instead. Thus, we are given a finite series of the form \( g(y) = \sum_{x \leq z \leq y} f(z) \) or \( g(x) = \sum_{x \leq z \leq y} f(z) \) where \( x \) is regarded as fixed in the former equation (giving \( g \) as a function of the variable \( y \)) and \( y \) is regarded as fixed in the latter. A given function \( f: P \to R \) may be regarded as a function in the incidence algebra in two natural ways. We may define \( f[x,y] = f(x) \) for all intervals \( x,y \) or we may define \( \hat{f}[x,y] = f(y) \). We call \( f \) the lower representation of \( f \) and \( \hat{f} \) the upper representation of \( f \). We then see that the identity \( g(y) = \sum_{x \leq z \leq y} f(z) \) is equivalent to the identity \( \hat{g} = \hat{f} \cdot \zeta \) and the identity \( g(x) = \sum_{x \leq z \leq y} f(z) \) is equivalent to the identity \( g = \xi \cdot f \). With these correspondences, LEMMA 5C.18 follows from LEMMA 5C.16.

**5C.18 LEMMA.**

Let \( f \) and \( g \) be functions from a poset \( P \) to \( R \). The following rules for inverting finite sums apply:

(1) If \( g(y) = \sum_{x \leq z \leq y} f(z) \) then \( f(y) = \sum_{x \leq z \leq y} g(z) \mu[z,y] \).

(2) If \( g(x) = \sum_{x \leq z \leq y} f(z) \) then \( f(x) = \sum_{x \leq z \leq y} \mu[x,z] g(z) \).

We shall now give some examples of the use of LEMMA 5C.18.
5C.19 EXERCISE.

(1) Prove LEMMA 5C.18.

(2) In LEMMA 5C.16 and LEMMA 5C.18, R denotes the real numbers. For what other algebraic structures are these lemmas valid? For example, if R were to denote all polynomials in a variable x, would these lemmas still be valid? Hint: This is an occasion where the "matrix point of view" might be helpful, at least conceptually.

5C.20 THE MÖBIUS FUNCTION OF THE POSET OF PARTITIONS OF A SET.

Consider the poset \( \Pi(D) \) of partitions of a finite set \( D \), \( |D| = d \), where \( \pi_1 \leq \pi_2 \) means "\( \pi_1 \) refines \( \pi_2 \)" (see DEFINITION 1.3 and EXAMPLE 1.9(4)). The "elements" of a partition are subsets of \( D \) and are called the "blocks" of the partition. We use the usual notation \(|\pi|\) to denote the number of blocks of the partition \( \pi \). Let \( S \) be a set with \(|S| = s\). For any function \( f \in S^D \) recall that the coimage(f) is the partition of \( D \) whose blocks are the sets \( f^{-1}(t) \) for \( t \in \text{image}(f) \) (see NOTATION 1.6). Given \( \pi \in \Pi(D) \), one may easily verify that \(|\{f: \text{coimage}(f) = \pi\}| = (s)_{|\pi|}\). Here, \((s)_k = s(s-1)(s-2) \ldots (s-k+1)\) is the falling factorial. Let \( \theta \) denote the discrete partition \( \{\{t\}: t \in D\} \) and let \( \tau \) denote the one block partition. By classifying the functions of \( S^D \) by the cardinality of their coimage, we obtain the identity \( s^d = s^{[0]} = \sum_{\theta \subseteq \pi \subseteq \tau} \mu(\theta, \pi)(s)^{|\pi|} \). Using the inversion formula of LEMMA 5C.18(2) we obtain \( (s)_d = (s)^{[0]} = \sum_{\theta \subseteq \pi \subseteq \tau} \mu(\theta, \pi)(s)^{|\pi|} \). This identity is valid for all values of \( s \), so we may view it as a polynomial identity and think of \( s \) as a variable. When we compare the coefficients of \( s \) on both sides of the identity, we obtain \( \mu[\theta, \tau] = (-1)^{d-1}(d-1)! \). Thus, we have computed the value of \( \mu \) on the total interval \([\theta, \tau]\).

Suppose now that we are given partitions \( \eta \leq \rho \in \Pi(D) \). Let \( B_1, B_2, \ldots, B_q \) denote the blocks of the partition \( \rho \). Since \( \eta \) is a refinement of \( \rho \), each block \( B_i \) is further partitioned by \( \eta \). We call this partition of \( B_i \) induced by \( \eta \), \( \eta_i \). Since \( \eta_i \) is a set (of blocks) we may consider \( \Pi(\eta_i) \). Note that the interval \([\eta_i, B_i]\) in the poset \( \Pi(B_i) \) is order isomorphic to the total interval in the poset \( \Pi(\eta_i) \). Hence, the interval \([\eta, \rho]\) in the poset \( \Pi(D) \) is isomorphic to the product \( \times \Pi(\eta_i) \). Thus, by THEOREM 5C.13, we obtain the formula \( \mu[\eta, \rho] = (-1)^{|\eta| - |\rho|} (|\eta_1| - 1)! (|\eta_2| - 1)! \ldots (|\eta_q| - 1)! \) where \( \eta_1, \ldots, \eta_q \) are the restrictions of \( \eta \) to the blocks of \( \rho \). This completely specifies the Möbius function of the poset of partitions of a set.

5C.21 MÖBIUS INVERSION AND THE PRINCIPLE OF INCLUSION-EXCLUSION.

In a certain sense, the principle of inclusion-exclusion can be regarded as a special case of Möbius inversion. Consider COROLLARY 5B.4. Let \( \mathcal{N}_-(T) = \Pi_{i \in T} f_i \Pi_{i \in T^c} (u - f_i) \) and let \( \mathcal{N}_+(T) = \Pi_{i \in T} f_i \). We observe that
\[
\sum_{T \subseteq R \subseteq n} \mathcal{N}_\simeq(R) = \sum_{T \subseteq R \subseteq n} \prod_{i \in R} f_i \prod_{i \notin R} (u-f_i) \\
= \prod_{i \in T} f_i \sum_{Q \subseteq R \setminus T} \prod_{i \in Q} f_i \prod_{i \notin (R \setminus T) \setminus Q} (u-f_i) \\
= \prod_{i \in T} f_i \prod_{i \notin R \setminus T} (f_i + (u-f_i)) = \prod_{i \in T} f_i = \mathcal{N}_\simeq(T).
\]

Thus we have \( \mathcal{N}_\simeq(T) = \sum_{T \subseteq R \subseteq n} \mathcal{N}_\simeq(R) \) which is a form of LEMMA 5C.18(2) where the poset is \( \mathcal{P}(n) \) (EXAMPLE 5C.14). We are actually using a slight extension of LEMMA 5C.18 suggested in EXERCISE 5C.19(2). Thus we see that

\[
\mathcal{N}_\simeq(T) = \sum_{T \subseteq R \subseteq n} \mu(T,R) \mathcal{N}_\simeq(R)
\]

when \( \mu(T,R) = (-1)^{|R|-|T|} \). This gives COROLLARY 5B.4 from which all other results on inclusion-exclusion follow. Of course, the proof of COROLLARY 5B.4 is a trivial algebraic manipulation, so using Möbius inversion to prove the result is a case of “using a sledgehammer to kill a fly.” We point out that we use \( P \) and \( P_i \) as sets in connection with the discussion of inclusion-exclusion and as posets in connection with Möbius inversion. In both cases this is standard notation. In the former case, \( P_i \) stands for the subset of \( P \) with “property” \( i \) (in the case of derangements, \( P_i \) is the subset of \( P = S_n \) with the property that \( i \) is a fixed point). In COROLLARY 5B.2, for example, the underlying poset is \( (\mathcal{P}(n), \subseteq) \). The \( P_i \) are just subsets of a set \( P \) and are not to be confused with the posets of the Möbius inversion discussion.

5C.22 MOBIUS INVERSION AND DIVISIBILITY OF INTEGERS: EULER’S \( \varphi \)-FUNCTION.

We refer to EXAMPLE 5C.15. For a positive integer \( n \), we let \( \varphi(n) \) denote the number of positive integers \( x \) that are less than or equal to \( n \) and are relatively prime to \( n \) (g.c.d.(\( x, n \)) = 1, where g.c.d. stands for “greatest common divisor”). The function \( \varphi \) is called the Euler \( \varphi \)-function. Let \( S_d = \{i \leq n: \text{g.c.d.}(i,n) = d\} \). It is evident that \( n = \bigcup_{d|n} S_d \) and the sets \( S_d \) are disjoint. Thus \( n = \sum_{d|n} |S_d| \). Note that \( |S_d| = \varphi(n/d) \) and hence \( n = \sum_{d|n} \varphi(n/d) = \sum_{d|n} \varphi(d) \). This identity is of the form of LEMMA 5C.18(1) so we obtain the identity \( \varphi(n) = \sum_{d|n} d \mu(d,n) \).

Now, let \( \{p_i: i \in T\} \) be the set of primes that are divisors of \( n \). From EXAMPLE 5C.15 we see that \( \mu(d,n) = 0 \) unless \( d = n/\prod_{i \in Q} p_i \) where \( Q \subseteq T \), in which case \( \mu(d,n) = (-1)^{|Q|} \). Thus, \( \varphi(n) = \sum_{Q \subseteq T} (-1)^{|Q|} n/\prod_{i \in Q} p_i = n \prod_{p|n} (1 - 1/p) \) where
the product is over all prime divisors of $n$. This result should be compared with EXAMPLE 5B.8.

5C.23 MöBIUS INVERSION AND THE POSET OF PARTITIONS: CONNECTED GRAPHS.

For this example, the reader should look at the material in CHAPTER 6 up to DEFINITION 6.11. Let $G = (V, E)$ be a graph (no loops). From the definition, it is immediate that there are $2\binom{n}{2}$ such graphs if $|V| = n$. Associated with each graph $G$ is a partition $\pi$ of $V$ determined by the connected components of $G$ (each block of $\pi$ is a connected component's vertex set). We call $\pi$ the component partition of $G$. We say that a partition $\pi$ is of type $(b_1, b_2, \ldots, b_n)$ if there are $b_i$ blocks of cardinality $t$ in $\pi$, $t = 1, \ldots, n$. We have seen this idea in connection with ranking ordered partitions with fixed multinomial index (see TABLE 3.28 and the related discussion). The idea of type was also discussed in connection with permutations (in the derivation of IDENTITY 4.46) and for partitions (in connection with FIGURE 4.63). One can show that the number of partitions $\pi$ with type $(b_1, b_2, \ldots, b_n)$ is $n! / \prod_{t=1}^{n} (t!)^{b_t} b_t!$ (EXERCISE 5C.24(1)). As in EXAMPLE 5C.20, we let $\theta$ be the discrete partition (minimal element of $\Pi(V)$) and $\tau$ the one-block partition (maximal element of $\Pi(V)$). For a given partition $\pi$, let $N_\leq(\pi)$ denote the number of graphs $G = (V, E)$ with component partition $\pi$ and let $N_\leq(\pi)$ denote all graphs whose component partition is a refinement of $\pi$. We thus have

$$2^{\binom{n}{2}} = N_\leq(\tau) = \sum_{\theta \leq \pi \leq \tau} N_\leq(\pi).$$

This identity is in the form of LEMMA 5C.18(1). Applying LEMMA 5C.18(1) we obtain

$$N_\leq(\tau) = \sum_{\theta \leq \pi \leq \tau} N_\leq(\pi) \mu(\pi, \tau)$$

where $\mu$ is the Möbius function of the poset $\Pi(V)$ (EXAMPLE 5C.20). Thus

$$\mu(\pi, \tau) = (-1)^{\Sigma b_i - 1} (\Sigma b_i - 1)!. It is easily seen that $N_\leq(\pi) = \prod_i \binom{j_i}{b_i}$. $N_\leq(\tau)$ is the number $c_n$ of connected graphs $G = (V, E)$ with $|V| = n$. Thus we obtain

$$c_n = \sum_{b} (-1)^{\Sigma b_i - 1} \left( \sum b_i - 1 \right)! \prod_i \binom{j_i}{b_i} \frac{n!}{\prod_i (i!)^{b_i} b_i!}$$
where the sum is over all possible types \((b_1, \ldots, b_n)\) or all \(n\)-tuples of non-negative integers \((b_1, \ldots, b_n)\) such that \(\sum_i ib_i = n\). Simplifying, we obtain that the number \(c_n\) of connected graphs \(G = (V,E)\) with \(|V| = n\) is:

\[
c_n = -n! \sum_b \left( \sum b_i - 1 \right)! \prod_i \left[ -2 \binom{i}{2} \right]^{b_i} \frac{1}{b_i!}
\]

where the product is over \(b_i > 0\) and the sum is over all \((b_1, \ldots, b_n)\) such that \(\sum_i ib_i = n\). Unfortunately, the number of terms in this sum grows exponentially with \(n\) so the formula is only useful for small values of \(n\).

5C.24 EXERCISE.

(1) Try to give a reasonably convincing argument that the number of partitions \(n\)
\[\pi\] with type \((b_1, b_2, \ldots, b_n)\) is \(n! \prod_{t=1}^n (t!)^{b_t} b_t!\). Hint: The argument is similar to that used in deriving IDENTITY 4.46.

(2) Additional applications of Möbius inversion (Waring’s formula, convex polytopes, finite vector space, etc.) are given in the references at the end of Part I. Organize and present some of this material to the class.

5D. NETWORK FLOWS.

For this section, the reader should study the material in CHAPTER 6 through DEFINITION 6.11. FIGURE 5D.1(a) shows a directed graph \(G = (V,E)\) with \(V = \{s,p,q,r,t\}\). Each edge is labeled with a value \(c\), called the “capacity of the edge” and a value \(f\), called the “flow of the edge.” The vertex \(s\) is called the “source” of \(G\) and the vertex \(t\) is called the “sink” of \(G\). Imagine that each edge represents a tube through which water is flowing in the direction indicated. The capacity \(c\) of each edge or “tube” represents the maximum amount of water that can flow through the tube in the indicated direction (in m\(^3\)/sec, say). Suppose that at all vertices except \(s\) and \(t\) the amount of water flowing into that vertex must equal the amount of water flowing out of that vertex. Generally, one thinks of water being put into the system at \(s\) and flowing out at \(t\), but the roles of \(s\) and \(t\) are symmetrical. The flow \(f\) represents the actual amount of water flowing through the “network.”
5D.1 FLOW THROUGH A NETWORK.

![Diagram of a network flow problem]

Figure 5D.1 (a)

The undirected path \((s,q,p,t)\) with edge sequence \((s,q),(q,p),(t,p)\) is augmentable. Edges \((s,q)\) and \((q,p)\) are forward directed, edge \((t,p)\) backward directed. The edge sequence \((s,q),(q,p),(p,t)\) also may be associated with this path.

![Diagram of a vertex with edges]

Figure 5D.1 (b)

\[
\text{OUTOF}(v) = e_1 + e_2 + e_4 + e_5 \\
\text{INTO}(v) = e_3 + e_6 \\
\text{STAR}(v) = e_1 + e_2 + e_4 + e_5 - e_3 - e_6 \\
\text{LINEAR EXTENSION for } h: E \rightarrow R, h(\text{STAR}(v)) = h(e_1) + h(e_2) + h(e_4) + h(e_5) - h(e_3) - h(e_6)
\]
Before proceeding further with networks, we discuss briefly an important conceptual idea. Let \( G = (V,E) \) be a directed graph. We think of the symbols \( e_1, e_2, \ldots \) representing the edges of \( G \) as variables. In this way we can form polynomials in these variables such as \( 4e_1 + e_3^2 \), etc. Actually, we shall only need linear polynomials such as \( e_1 - 2e_2 + 3e_3 \), etc. The set of all linear polynomials with \( E \) as a set of variables and real numbers as coefficients has a fancy mathematical name: \textit{the free vector space of} \( E \) \textit{over the real numbers}.

Name aside, the idea is trivial but notationally and conceptually useful. If \( h: E \to \mathbb{R} \) is any function from \( E \) to the real numbers then we define the value of \( h \) on a polynomial \( p(e_1, \ldots, e_n) \) by \( hp(e_1, \ldots, e_n) = p(h(e_1), \ldots, h(e_n)) \). For example, \( h(e_1 - 3e_2 + 4e_3 - 2e_4) = h(e_1) - 3h(e_2) + 4h(e_3) - 2h(e_4) \). This extension of \( h \) to linear polynomials is called “linear extension” of \( h \). In Definition 5D.2, we specify some basic polynomials associated with a directed graph.

5D.2 Definition.

Let \( G = (V,E) \) be a directed graph without loops (i.e., no edges of the form \((x,x)\)). For each \( v \) in \( V \), let \( \text{INTO}(v) \) be the polynomial formed by summing all edges incident on \( v \) and directed towards \( v \), let \( \text{OUTOF}(v) \) be the sum of all edges incident on \( v \) and directed away from \( v \), and let \( \text{STAR}(v) = \text{OUTOF}(v) - \text{INTO}(v) \). An example is shown in Figure 5D.1(b).

Referring again to Figure 5D.1, we give the formal definitions needed in connection with networks in Definition 5D.3.

5D.3 Definition.

A network is a 4-tuple \((G,s,t,c)\) where \( G \) is a directed graph, \( s \) and \( t \) are distinct vertices of \( G \) called the source and the sink respectively, and \( c: E \to \mathbb{R}^+ \) is a function from the edges \( E \) of \( G \) to the positive real numbers. The function \( c \) is called the capacity of the network.

In Definition 5D.4, we define the “flow” of a network in terms of the polynomials defined in Definition 5D.2.

5D.4 Definition.

Let \((G,s,t,c)\) be a network. A function \( f: E \to \mathbb{R}^+_0 \) that maps \( E \) to the non-negative real numbers and satisfies \( f(\text{STAR}(v)) = 0 \) for all \( v \) not equal to \( s \) or \( t \) and \( f(e) \leq c(e) \) for all \( e \in E \) will be called a flow on the network.

The intuitive idea of a flow has been presented above in connection with water flowing in tubes (Figure 5D.1). The condition \( f(\text{STAR}(v)) = 0 \) says that “the water flowing into \( v \) equals the water flowing out of \( v \).”
5D.6 Lemma

For each edge \( e \in E \), let \( \text{IN}(e) \) denote the initial vertex of \( e \) and let \( \text{TM}(e) \) denote the terminal vertex of \( e \). (See NOTATION 6.41 of CHAPTER 6 for more discussion.)

5D.5 DEFINITION.

Let \( (G,s,t,c) \) be a network. A partition \( \{P,\overline{P}\} \) of \( V \) into two blocks (DEFINITION 1.3) with \( s \in P \) and \( t \in \overline{P} \) will be called a cut partition. Let \( \text{OUTCUT}(P,\overline{P}) \) be the polynomial formed by summing all edges \( e \) with \( \text{IN}(e) \in P \) and \( \text{TM}(e) \in \overline{P} \) (i.e., all edges that go from \( P \) to \( \overline{P} \)) and let \( \text{INCUT}(P,\overline{P}) \) be the sum over all edges going from \( \overline{P} \) to \( P \). Define \( \text{CUT}(P,\overline{P}) = \text{OUTCUT}(P,\overline{P}) + \text{INCUT}(P,\overline{P}) \).

We define the number \( c(\text{OUTCUT}(P,\overline{P})) \) to be the capacity of the cut partition \( \{P,\overline{P}\} \), or the capacity of the cut.

In terms of the "water in tubes" analogy, the capacity of \( \{P,\overline{P}\} \) represents the maximum amount of water that could possibly be flowing from \( P \) to \( \overline{P} \).

5D.6 LEMMA.

Let \( (G,s,t,c) \) be a network and let \( \{P,\overline{P}\} \) be a partition of \( V \) with \( s \in P \) and \( t \in \overline{P} \) as in DEFINITION 5D.5. Using the notation of DEFINITIONS 5D.2 and 5D.5, the following identities hold:

\[
\begin{align*}
(1) \quad & \sum_{v \in P} \text{STAR}(v) = \text{OUTCUT}(P,\overline{P}) - \text{INCUT}(P,\overline{P}). \\
(2) \quad & \sum_{v \in \overline{P}} \text{STAR}(v) = \text{INCUT}(P,\overline{P}) - \text{OUTCUT}(P,\overline{P}). \\
(3) \quad & \sum_{v \in P} \text{STAR}(v) = -\sum_{v \in \overline{P}} \text{STAR}(v). \\
(4) \quad & \text{For any flow } f, f(\text{STAR}(s)) = -f(\text{STAR}(t)) = f(\text{OUTCUT}(P,\overline{P})) - f(\text{INCUT}(P,\overline{P})). \text{ This number is called the value of the flow.} \\
(5) \quad & \text{For any flow } f, f(\text{STAR}(s)) \leq c(\text{OUTCUT}(P,\overline{P})) \text{ (the value of the flow is less than or equal to the capacity of the cut).}
\end{align*}
\]

Proof. We first prove (1). Let \( e = (v,w) \) be an edge with both \( v \) and \( w \) in \( P \). Then, \( e \) occurs in the polynomial \( \text{STAR}(v) \) and \( -e \) occurs in the polynomial \( \text{STAR}(w) \). The edge \( e \) does not occur in \( \text{STAR}(x) \) if \( x \) is not equal to \( v \) or \( w \). Thus all edges with both vertices in \( P \) cancel in forming the sum that is the left-hand side of identity (1). This leaves only the right-hand side of identity (1). Identity (2) is derived in the same manner. Identity (3) follows trivially from (1) and (2). To prove (4), note that, from (1), \( \sum_{v \in P} f(\text{STAR}(v)) = f(\text{OUTCUT}(P,\overline{P})) \)
- \( f(\text{IN Cut}(P, \overline{P})) \) for any function \( f \), whether or not \( f \) is a flow. If \( f \) is flow, however, we have \( f(\text{Star}(v)) = 0 \) if \( v \neq s \) or \( t \). As \( s \in P \) but \( t \notin P \), identity (4) follows. To prove (5), note that the stronger identity \( f(\text{Star}(s)) \leq f(\text{Out Cut}(P, \overline{P})) \) is obvious from (4). From the definition of a flow, \( f(\text{Out Cut}(P, \overline{P})) \leq c(\text{Out Cut}(P, \overline{P})) \).

Note that the inequality of LEMMA 5D.6(5) is valid for any flow and any partition \( \{P, \overline{P}\} \) such that \( s \in P \) and \( t \notin \overline{P} \). In particular, the maximum value over all flows \( f \) of \( f(\text{Star}(s)) \) is less than or equal to the minimum over all such \( \{P, \overline{P}\} \) of \( c(\text{Out Cut}(P, \overline{P})) \). In fact, we shall see by THEOREM 5D.8 that these numbers are always equal.

5D.7. DEFINITION.

Let \((G, s, t, c)\) be a network. The maximum value over all flows \( f \) on the network of \( f(\text{Star}(s)) \) will be denoted by MAXFLOW. The minimum over all partitions \( \{P, \overline{P}\} \) of \( c(\text{Out Cut}(P, \overline{P})) \) will be denoted by MINCUT.

As we have remarked, the fact that MAXFLOW \( \leq \) MINCUT follows easily from LEMMA 5D.6(5).

5D.8 THEOREM.

For any network \((G, s, t, c)\), MAXFLOW = MINCUT.

Proof. Let \( f \) be any flow. Let \( \tilde{G} \) be the undirected version of the graph \( G \). Consider a path \( s = x_1, x_2, \ldots, x_j = v \) in \( \tilde{G} \) from \( s \) to a vertex \( v \). A directed edge of \( G \) of the form \( (x_i, x_{i+1}) \), \( 1 \leq i < j \), will be called forward directed relative to this path and an edge of \( G \) of the form \( (x_{i+1}, x_i) \) backward directed. If it is possible to choose a sequence of such edges \( (e_1, e_2, \ldots, e_{j-1}) \) of \( G \) such that for each forward directed \( e_i = (x_i, x_{i+1}) \), \( f(e_i) < c(e_i) \) and for each backward directed \( e_i = (x_{i+1}, x_i) \), \( f(e_i) > 0 \) then the path \( x_1, \ldots, x_j \) will be called augmentable relative to the flow \( f \). The sequence \( (e_1, e_2, \ldots, e_{j-1}) \) will be called an edge sequence of the path, and may not be unique (see FIGURE 5D.1(a)).

Let \( P_f \) denote the set of all vertices \( v \) of \( G \) for which there exists an augmentable path from \( s \) to \( v \). We assume \( s \in P_f \). Suppose that \( t \notin P_f \) and let \( s = x_1, x_2, \ldots, x_k = t \) be an augmentable path joining \( s \) to \( t \) with edge sequence \( (e_1, e_2, \ldots, e_{k-1}) \). Let \( M \) be the set of all numbers of the form \( c(e) - f(e) \) for "forward-directed" edges \( e = (x_i, x_{i+1}) \) and of the form \( f(e) \) for "backward-directed" edges \( e = (x_{i+1}, x_i) \). Let \( m > 0 \) be the minimum value of \( M \). Define a new flow \( f' \) on \( G \) by \( f'(e) = f(e) + m \) on each forward-directed edge of the augmentable path, \( f'(e) = f(e) - m \) on each backward-directed edge of the augmentable path, and \( f'(e) = f(e) \) for all other edges of \( G \). It is easily seen that \( f' \) is a flow and that \( f'(\text{Star}(s)) > f(\text{Star}(s)) \) (i.e., the value of \( f' \) is greater than the value of \( f \) so \( f \) cannot be a flow of maximal value). By elementary topological considerations
(trivial in the case of "integral flows" where $f$ and $c$ take on only non-negative integral values) there exists a flow $F$ of maximal value. By what we have just shown, $t \notin P_F$ so $\{P_F, \overline{P_F}\}$ is a cut partition (DEFINITION 5D.5). By the definition of $P_F$, any edge $e$ in OUTCUT($P_F, \overline{P_F}$) must have $F(e) = c(e)$ and any edge in INCUT($P_F, \overline{P_F}$) must have $F(e) = 0$. Thus, by LEMMA 5D.6(4), $F($STAR$(s)) = c($OUTCUT($P_F, \overline{P_F}$)). This completes the proof as we remarked above, MAX-FLOW $\leq$ MINCUT is a trivial consequence of LEMMA 5D.6.

There are many applications of THEOREM 5D.8. We shall mention some of the more interesting combinatorial consequences. The literature on this subject is extensive and quite accessible. The classical reference, and still one of the best, is the book *Flows in Networks* by Ford and Fulkerson (Princeton University Press, 1962).

An important technical result is stated in THEOREM 5D.9.

5D.9 **THEOREM** (Integrity theorem).

For any network $(G, s, t, c)$ where $c$ is integral valued there exists a maximal flow $f$ that is integral valued.

**Proof.** An inspection of the proof of THEOREM 5D.8 reveals that the augmentation process can always be done in such a way as to transform an integral valued flow into another integral valued flow of larger value.

One obvious and important extension of THEOREM 5D.8 is to networks of the form $(G, s, t, c, d)$ where $(G, s, t, c)$ is a network as above but $d: V \rightarrow R^+$ is a function from the vertices $V$ of $G$ to the positive real numbers (a vertex capacity function). We must have $f($INTO$(v)) \leq d(v)$ and $f($OUTOF$(v)) \leq d(v)$ for all $v$. A simple trick of replacing each vertex $v \in V$ by a directed edge reduces such networks to the case where only an edge capacity is considered. This idea is shown in FIGURE 5D.10. In the second network of FIGURE 5D.10, the source is $s$ and the sink is $t'$. 
5D.10 VERTEX CAPACITIES REDUCED TO EDGE CAPACITIES.

Another trivial extension of THEOREM 5D.8 arises when the source s is replaced by a set S of sources and the sink t is replaced by a set T of sinks. Thus one has a network of the form (G, S, T, c). Such networks can be reduced to the case of a single source and sink as indicated in FIGURE 5D.11 where a single source s and sink t are added to the network. An edge of infinite (the same as "finite but very large" in this case) capacity is added to the network joining s to each element of S and each element of T to t.

5D.11 MANY SOURCES AND MANY SINKS.

An important graph theoretic version of THEOREM 5D.8 is called "Menger's Theorem" (THEOREM 5D.12). Let $G' = (V', E')$ be an undirected graph and
let $S'$ and $T'$ be two disjoint subsets of $V'$. Let $a_1, a_2, \ldots, a_n$ be the vertex sequence of a path in $G'$ (see CHAPTER 6, DEFINITION 6.9). The vertex $a_1$ is called the "initial vertex" of the path and the vertex $a_n$ is called the "terminal vertex." The other vertices are called "internal vertices" of the path. Two paths in $G'$ will be called vertex disjoint if they have no vertices in common and internally vertex disjoint if they have no internal vertices in common. Let $S'$ and $T'$ be two nonempty disjoint subsets of $V'$. A subset $D'$ of $V'$ will be called an $S', T'$ blocking set (or disconnecting set, or separator) if every path with initial vertex in $S'$ and terminal vertex in $T'$ (i.e., a path from $S'$ to $T'$) has a vertex in $D'$ (passes through $D'$). In what follows, "a path from $S'$ to $T'$" means a path with initial vertex in $S'$, terminal vertex in $T'$ and all other vertices in $V' - S' - T'$.

5D.12 THEOREM (Menger).

Let $G' = (V', E')$ be an undirected graph and let $S'$ and $T'$ be nonempty disjoint subsets of $V'$. The maximum number of pairwise vertex disjoint paths from $S'$ to $T'$ is equal to the minimum size of an $S', T'$ blocking set $D'$.

Proof. (See EXERCISE 5D.21(1).)

A frequently occurring special case of THEOREM 5D.12 happens when $S' \cup T' = V'$ and all edges of $G'$ have one vertex in $S'$ and the other in $T'$. In this case, the graph is called bipartite with vertex partition $\{S', T'\}$. An example of a bipartite graph is shown in FIGURE 5D.14. In this case the only paths starting in $S'$ and ending in $T'$ are single edges. Any $S', T'$ blocking set $D'$ must be a subset of $S' \cup T'$ as this set is all of $V'$. In this case, the set $D'$ is sometimes referred to as a "vertex cover" of the bipartite graph because the union of all edges of $G'$ incident on $D'$ is all of $E'$ ("covers" $E'$). A picturesque rephrasing of Menger's theorem in the bipartite case is sometimes called the "rook-plank" theorem (EXERCISE 5D.21(2)).

5D.13 COROLLARY.

Let $G'$ be a bipartite graph with vertex partition $\{S', T'\}$. The maximum size of a pairwise disjoint collection of edges is equal to the minimum size of a vertex cover.

If $G'$ is a bipartite graph with vertex partition $\{S', T'\}$ then each edge $e = \{x, y\}$, $x \in S'$, $y \in T'$, may be thought of as "matching some element of $S'$ to an element of $T'$." Correspondingly, a collection of disjoint edges may be thought of as matching a subset of $S'$ to a subset of $T'$ (i.e., defining a bijection between these subsets). Thus we call any collection of disjoint edges of $G'$ a matching. Suppose $|S'| \leq |T'|$. A subset of edges of $G'$ that is pairwise disjoint (i.e., is a matching) and whose union contains $S'$ is called a complete matching of $S'$. 
Conditions for the existence of complete matchings are sometimes of interest. One standard example is the case where $S'$ is a set of girls and $T'$ is a set of boys. A bipartite graph is constructed by defining $\{x,y\}$ to be an edge if girl $x$ knows boy $y$. A dating service might be interested in a complete matching that would pair each girl with one boy that she already knows. FIGURE 5D.14 shows a bipartite graph with a complete matching.

5D.14 A BIPARTITE GRAPH WITH A COMPLETE MATCHING.

\[
\begin{array}{c}
\text{SOLID EDGES GIVE A COMPLETE MATCHING} \\
\text{Figure 5D.14}
\end{array}
\]

Let $G'$ be a graph and let $x$ be a vertex of $G'$. A vertex $y$ for which there is an edge of the form $\{x,y\}$ is said to be adjacent to $x$. Let $A_x$ denote the set of all vertices adjacent to $x$. THEOREM 5D.15 gives the standard conditions for the existence of a complete matching in a bipartite graph.

5D.15 THEOREM (Matching theorem).

Let $G'$ be a bipartite graph with vertex partition $\{S',T'\}$. There exists a complete matching of $S'$ if and only if for every subset $Q \subseteq S'$, \[ |\bigcup_{x \in Q} A_x| \geq |Q|. \]

Proof. The “only if” part is trivial. Suppose the inequality of the hypothesis holds. Consider a minimal vertex cover $D' = A' \cup B'$ where $A' \subseteq S'$ and $B' \subseteq T'$. The situation is shown in FIGURE 5D.16. Note that by the definition of $D'$ there can be no edge from $S' - A'$ to $T' - B'$. Thus $B' \supseteq \bigcup_{x \in S' - A'} A_x$ and, by our hypothesis, $|B'| \geq |\bigcup_{x \in S' - A'} A_x| \geq |S' - A'|$. Thus $|D'| = |A'| + |B'| \geq |S'|$. Of course $|D'| = |S'|$ as $S'$ is itself a vertex cover. The result now follows immediately from COROLLARY 5D.13.
5D.16 MINIMAL VERTEX COVER FOR MATCHING THEOREM.

\[ B' \quad T' \quad B' \]
\[ S' \quad A' \quad S' - A' \]

\[ \text{NO SUCH EDGES POSSIBLE} \]

\[ D' = A' \cup B' \text{ IS THE MINIMAL VERTEX COVER} \]

Figure 5D.16

As our final example, we consider Dilworth's theorem for ordered sets (DEFINITION 1.8). An ordered set \( P = (X, \alpha) \) is also called a "poset" and we say "\( x \) is an element of \( P \)" when \( x \in X \). A subset \( Q \subseteq X \) which has the property that \( x, y \in Q \) implies that \( x \prec y \) or \( y \prec x \) is called a linearly ordered subset of \( P \) or a chain. A partition of \( P \) in which each block is a chain is called a "chain cover" of \( P \). A subset \( T \subseteq Q \) with the property that if \( x, y \in T \) then \( x \prec y \) implies that \( x = y \) is called a transversal for \( P \). Note that any single element \( x \in P \) defines a singleton chain \( \{x\} \). The cardinality of a chain cover \( \mathcal{C} \) is the number of blocks in \( \mathcal{C} \). We say that "\( s \) and \( t \) are comparable in \( P \)" if either \( s \prec t \) or \( t \prec s \). A transversal \( T \) is a subset of \( X \) such that every pair of distinct elements is not comparable. The length of a chain \( Q \) is \( |Q| - 1 \).

5D.17 THEOREM (Dilworth).

The minimum cardinality of a chain cover of a poset \( P \) equals the maximum cardinality of a transversal of \( P \).

Proof. We use COROLLARY 5D.13. For convenience of notation, let \( P = (n, \alpha) \) and let \( P' = (n', \alpha) \) be an order isomorphic copy of \( P \) where \( n' = \{1', 2', \ldots, n'\} \). Construct a bipartite graph \( G' \) with vertex partition \( S' = n \) and \( T' = n' \) by defining a pair \( \{i, j'\} \) to be an edge if \( i \prec j \) in the poset \( P \). An example is shown in FIGURE 5D.18. The poset there is specified by its Hasse diagram (DEFINITION 1.10).
5D.18 BIPARTITE GRAPH OF A POSET.

We now define a bijection $\beta$ between disjoint edge sets of $G'$ and chain covers of $P$. Suppose we are given a disjoint edge set, DES, in $G'$. Every edge in DES corresponds to a 2-vertex chain in $P$, e.g., $\{1,3'\} \leftrightarrow \{2\}$. If any two of the chains thus formed from DES are not disjoint, then they must be of the form $\{a\}^b$ and $\{b\}^d$ (by disjointness of their corresponding edges in $G'$). We can link these chains to form $\{a\}^d$. By forming all such linkages we construct a set of disjoint chains in $P$. Let $\beta(\text{DES}) = \{\text{disjoint chains in } P \text{ as constructed above}\} \cup \{\text{all points left in } P, \text{considered as singleton chains}\}$. Clearly, $\beta(\text{DES})$ is a chain cover of $P$. Note that $\beta$ is an injection: if DES and DES' are two different disjoint edge sets of $G'$, $\beta(\text{DES})$ and $\beta(\text{DES}')$ are two distinct chain covers for $P$. To see that $\beta$ is a bijection, consider a chain cover CHC of $P$. Break every (nonsingleton) chain of length $n$ in $P$ into $n$ (nondisjoint) 2-chains. ($n \geq 1$), e.g.

\[
\begin{align*}
\{ & e \\
\{ & d \\
\{ & b \rightarrow \}^b_a \\
\{ & a \\
\end{align*}
\]

\[
\begin{align*}
\{ & d \\
\{ & b \\
\{ & e \\
\{ & d \\
\}
\end{align*}
\]

Associate with each 2-chain an edge in $G'$:

\[
\begin{align*}
\{ & b \\
\{ & a \leftarrow (a,b').
\end{align*}
\]

Note that by the way we formed the 2-chains, their corresponding edges in $G'$ will be disjoint. Let $\gamma(\text{CHC})$ be the disjoint edge set formed in this manner. Clearly, for any disjoint edge set DES,
\( \gamma(\beta(\text{DES})) = \text{DES} \)

and for any chain cover CHC in P,

\( \beta(\gamma(\text{CHC})) = \text{CHC} \).

So \( \gamma = \beta^{-1} \). \( \beta \) as defined above is a bijection.

Now, consider any disjoint edge set \( \text{DES} \) in \( G' \) and look at the chain cover \( \beta(\text{DES}) \) in P. Circle the top element in each of the chains in \( \beta(\text{DES}) \), and circle all singleton-chain points. (See FIGURE 5D.19.) Clearly the number of circled points is equal to \( |\beta(\text{DES})| \). Notice that to each chain of length \( k \) (\( k \geq 1 \)) in \( \beta(\text{DES}) \), there correspond \( k \) edges in \( \text{DES} \). Also, each chain of length \( k \) has \( k \) uncircled points. So the number of uncircled points is equal to \( |\text{DES}| \). But the number of circled points plus the number of uncircled points equals \( |P| \). Hence \( |\text{DES}| + |\beta(\text{DES})| = |P| \).

5D.19 RELATING \( |\beta(\text{DES})| \) TO \( |\text{DES}| \).

The above ideas relate disjoint edge sets and chain covers. We now must relate vertex covers and transversals. Consider any vertex cover \( \text{VC} \) of \( G' \). Let \( A = \text{VC} \cap S' \) and let \( B' = \text{VC} \cap T' \). Let \( B \) be the copy of \( B' \) in \( S' \) and let \( A' \) be the copy of \( A \) in \( T' \). These sets are shown in FIGURE 5D.20.
5D.20 RELATING TRANSVERSALS TO VERTEX COVERS.

\[ A' \quad W' \quad B' = T' \cap VC \]

\[ A = S' \cap VC \quad W \quad B \]

\[ A \cup B' \text{ is a vertex cover for } G' \]
\[ W \text{ is a transversal for } P \]

Figure 5D.20

Notice that we cannot have an edge going from \( W = S' - A - B \) to \( W' = T' - A' - B' \) (the copy of \( S' - A - B \)), as this would violate the fact that \( A \cup B' \) is a vertex cover. But this is equivalent to saying that no two points in \( S' - A - B \) are comparable (i.e., \( S' - A - B \) is a transversal). Hence, given any vertex cover \( VC \), we can define \( A \) and \( B \) as above, and let \( \tau(VC) \) equal the transversal \( W = S' - A - B \). Conversely, given a transversal \( W \), we can construct a VC such that \( \tau(VC) = W \). Note that, for any vertex cover \( VC \), \( |VC| + |\tau(VC)| \geq |P| \) (\( A \cap B \neq \emptyset \) is possible).

Given the two basic identities \( |DES| + |\beta(DES)| = |P| \) and \( |VC| + |\tau(VC)| \geq |P| \), Dilworth’s theorem follows easily, for maximizing the size of a disjoint edge set, \( |DES| \), corresponds to minimizing the size of a chain cover, \( |\beta(DES)| \).

By COROLLARY 5D.13, the maximum \( |DES| \) is equal to the minimum \( |VC| \), and thus, by the two basic identities above, the maximum possible \( |\tau(VC)| \) is greater than or equal to the minimum possible \( |\beta(DES)| \). The reverse inequality is trivial (each chain contains at most one transversal element).

5D.21 EXERCISE.

(1) Prove Menger’s theorem (THEOREM 5D.12). Hint: Construct a network with \( S' \) as the set of sources and \( T' \) as the set of sinks by replacing each edge of \( G' \) by a pair of opposite directed edges. Put unit capacity on all vertices and infinite capacity on all edges. Use the integrity theorem (THEOREM 5D.9).

(2) Consider an \( n \times n \) chessboard diagram with rows labeled \( r_1, r_2, \ldots, r_n \) and columns labeled \( c_1, c_2, \ldots, c_n \). Draw \( N \) rooks on the chessboard. Let NAT denote the size of the largest subset of these rooks such that no two rooks attack each other. Suppose that one has available a box of \( n \times 1 \) planks (very thin but not transparent) that can be used to cover a row or column of the board. Let PLK denote the smallest number of planks that can be laid on
the board (each plank covering a row or a column) such that all rooks are covered. The statement that $\text{NAT} = \text{PLK}$ is just COROLLARY 5D.13 in disguise. Can you explain why?

(3) Given a network $(G,s,t,c)$ how would you actually find a maximum flow and a minimum cut partition? A careful look at the proof of THEOREM 5D.8 gives one basic idea in the form of the "augmentable path." Try to devise your own algorithm for finding a maximum flow in a network, and illustrate it with some examples. Try to avoid looking up the standard algorithms given in the references until you have at least made a serious attempt yourself!

(4) State and prove a version of Menger's theorem for directed graphs. Hint: Menger's theorem in its various forms can be found in the references if you have difficulty with this problem or become interested in learning more about this class of results.