Advances in Mathematics **138**, 293–305 (1998)
Article No. AI981748

The Combinatorics behind Number-Theoretic Sieves

Timothy Y. Chow

Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109-1109
E-mail: tchow@alum.mit.edu

Received July 25, 1997; accepted March 2, 1998

Ever since Viggo Brun's pioneering work, number theorists have developed increasingly sophisticated refinements of the sieve of Eratosthenes to attack problems such as the twin prime conjecture and Goldbach's conjecture. Ever since Gian-Carlo Rota's pioneering work, combinatorialists have found more and more areas of combinatorics where sieve methods (or Möbius inversion) are applicable. Unfortunately, these two developments have proceeded largely independently of each other even though they are closely related. This paper begins the process of bridging the gap between them by showing that much of the theory behind the number-theoretic refinements carries over readily to many combinatorial settings. The hope is that this will result in new approaches to and more powerful tools for sieve problems in combinatorics such as the computation of chromatic polynomials, the enumeration of permutations with restricted position, and the enumeration of regions in hyperplane arrangements.

1. INTRODUCTION

The Möbius function of a finite partially ordered set has been a pervasive theme in combinatorics ever since Rota's revolutionary paper "Foundations I" [21]. Its ubiquity is quite astonishing; it is related to such diverse topics as the four-color theorem, the homology of simplicial complexes, and symmetric functions. Unexpected new applications are still being found today, e.g., Athanasiadis's finite field method for subspace arrangements [1] or Wagner's description of the coefficients of the Tutte polynomial of a matroid [24].

Rota was motivated in part by number theory, as is obvious from his choice of terminology (e.g., "Galois connection," or indeed the term "Möbius function" itself). So in view of the enormous success of his ideas, it is rather surprising that since Foundations I, there has been very little traffic carrying the deeper sieve methods developed by number theorists over into combinatorics. The purpose of this paper is to start this traffic rolling.

More specifically, we give a definition of a sieve that is sufficiently general to encompass many important sieve problems in both number

0001-8708/98 $25.00

theory and combinatorics and that is also sufficiently specialized to enable fairly sophisticated methods (such as Selberg's Λ^2 sieve) to be used. We also discuss briefly some potential applications.

2. SIEVE PROBLEMS

The term "sieve" is used frequently in both combinatorics and number theory, but there is no universally accepted technical definition of the term. We propose the following definition, which, while simple, is perhaps the most crucial idea in this paper.

DEFINITION. A *sieve* is an ordered 3-tuple $(\mathscr{A}, L, \sigma)$ where $\mathscr{A}$ is a finite set, L is a finite lattice, and σ is a map from $\mathscr{A}$ into L. A *weighted sieve* is an ordered 4-tuple $(\mathscr{A}, w, L, \sigma)$ where $\mathscr{A}$ is an arbitrary set, w is a function from $\mathscr{A}$ to the nonnegative reals such that $\sum_{a \in \mathscr{A}} w(a)$ converges, L is again a finite lattice, and σ is again a map from $\mathscr{A}$ into L.

As will become apparent in the discussion below, a sieve is equivalent to the special case of a weighted sieve in which the weight function equals one for a finite number of elements of $\mathscr{A}$ and equals zero everywhere else. The most interesting applications involve only the simpler concept of a sieve, but occasionally the extra generality of a weighted sieve is necessary. For this reason we will state our results in terms of weighted sieves.

The following notation will be used throughout this paper. Since L is finite, it has a minimum element $\hat{0}$ and a maximum element $\hat{1}$. We use $\prec$, $\wedge$, $\vee$, and μ to denote the partial ordering, the meet, the join, and the Möbius function of L respectively. If L is graded, we use ρ to denote its rank function. Basic facts about lattices may be found in [23, Chapter 3].

Given a weighted sieve $(\mathscr{A}, w, L, \sigma)$, define a function $A: L \to \mathbb{R}$ by

$$A(x) \stackrel{\text{def}}{=} \sum_{\{a \in \mathscr{A} : x \preccurlyeq \sigma(a)\}} w(a).$$

Also let $S: L \to \mathbb{R}$ be the function obtained from A by Möbius inversion:

$$S(x) \stackrel{\text{def}}{=} \sum_{y \succcurlyeq x} \mu(x, y) A(y).$$

PROPOSITION 1. *With the above terminology,*

$$S(\hat{0}) = \sum_{\{a \in \mathscr{A} : \sigma(a) = \hat{0}\}} w(a).$$

Proof. Define the *sifting function* $s_0: \mathscr{A} \to \{0, 1\}$ by

$$s_0(a) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if } \sigma(a) = \hat{0}; \\ 0, & \text{otherwise.} \end{cases}$$

Note that

$$s_0(a) = \sum_{y \leqslant \sigma(a)} \mu(\hat{0}, y).$$

Therefore

$$\begin{aligned} S(\hat{0}) &= \sum_{y \in L} \mu(\hat{0}, y)\, A(y) \\ &= \sum_{y \in L} \mu(\hat{0}, y) \sum_{\{a \in \mathscr{A} : y \leqslant \sigma(a)\}} w(a) \\ &= \sum_{a \in \mathscr{A}} w(a) \sum_{y \leqslant \sigma(a)} \mu(\hat{0}, y) \\ &= \sum_{a \in A} w(a)\, s_0(a) \\ &= \sum_{\{a \in \mathscr{A} : \sigma(a) = \hat{0}\}} w(a). \quad \blacksquare \end{aligned}$$

In a typical sieve problem, the function A is known (or at least approximately known) and the goal is to estimate $S(\hat{0})$. We may think of σ as distributing the elements of $\mathscr{A}$ over the elements of L; if we know A then we know how many elements of $\mathscr{A}$ lie *on or above* each element $x \in L$. So in light of Proposition 1, we see that wanting to know $S(\hat{0})$ is equivalent to wanting to know how many elements of $\mathscr{A}$ lie on $\hat{0}$.

Let us now look at some important examples, which we hope will clarify the above concepts and justify our choice of definitions. In each case, the weight function w will be identically one.

EXAMPLE 1. In a typical number-theoretic sieve problem, $\mathscr{A}$ is a set of the form

$$\mathscr{A} = \{h(1), h(2), ..., h(n)\}$$

where h is some integer-valued polynomial, L is the set of divisors (partially ordered by divisibility) of some fixed squarefree number Π, and $\sigma(a) = gcd(a, \Pi)$. Thus $A(x)$ is the number of elements of $\mathscr{A}$ that are

divisible by x, and by Proposition 1, $S(\hat{0})$ is the number of elements of $\mathscr{A}$ relatively prime to Π.

For instance, let h be the polynomial $h(r) = r(r+2)$, and let Π be the product of all primes less than or equal to $\sqrt{n} = |\mathscr{A}|^{1/2}$. Then $S(\hat{0})$ is the number of integers between 1 and n such that neither n nor $n+2$ is divisible by a prime less than or equal to $\sqrt{n}$. In other words, $S(\hat{0})$ is a good approximation to the number of twin primes less than or equal to n. Showing that $S(\hat{0}) \to \infty$ as $n \to \infty$ would prove the twin prime conjecture. Similarly, if h is the polynomial $h(r) = r(n-r)$ and Π is as before, then a sufficiently large lower bound for $S(\hat{0})$ would prove the Goldbach conjecture.

EXAMPLE 2. Let G be a finite abelian group, and let $\{H_1, H_2, ..., H_n\}$ be a family of subgroups of G. Let L be the lattice of all subgroups of G that are generated by some finite subset of the H's. Partially order L by inclusion. For some fixed integer k, let $\mathscr{A}$ be the set of all k-tuples $\chi = (\chi_1, \chi_2, ..., \chi_k)$ of characters of G. Define $\sigma(\chi)$ to be the *kernel* of χ, that is, the maximal element $K \in L$ such that χ_i restricted to K is trivial for all i. Then for any $H \in L$, $A(H)$ is the number of k-tuples of characters of G/H, i.e., $A(H) = (|G|/|H|)^k$. By Proposition 1, $S(\hat{0})$ is the number of k-tuples with trivial kernel.

Example 2 was first considered by Kung, Murty, and Rota [15, Theorem 10]. Its origins lie in Minkowski's conjecture that in any lattice tiling of $\mathbb{R}^n$ there exist two tiles that share an $(n-1)$-face. Hajós [11] solved this problem in 1942 by transforming it into a problem about abelian groups, and his work was extended by Rédei [18, 19]. The above observation about $S(\hat{0})$ yields a transparent proof of one of Rédei's main theorems (his so-called "Trägheitsatz").

It may not be clear why we care about the number of k-tuples with trivial kernel, so let us now consider two special cases of Example 2 where $S(\hat{0})$ counts something more obviously interesting. The first special case is a well-known result by Crapo and Rota on the so-called *critical problem* for finite vector spaces (see for example [3]).

EXAMPLE 2a. Take G to be an n-dimensional vector space $\mathbb{F}_q^n$ over the finite field $\mathbb{F}_q$, and let $\{v_1, v_2, ..., v_n\}$ be a set of vectors in $\mathbb{F}_q^n$. Let L be the lattice of all subspaces of $\mathbb{F}_q^n$ that are spanned by some finite subset of the v's. Characters correspond to linear functionals, so let $\mathscr{A}$ be the set of all k-tuples of linear functionals on $\mathbb{F}_q^n$. Define σ to be the kernel as before. Then it is not hard to see that

$$A(x) = q^{k(n-p(x))}$$

where $p(x)$ is the rank of x in L. As before, $S(\hat{0})$ is the number of k-tuples with trivial kernel.

The *characteristic polynomial* $\chi_L(\lambda)$ of L is defined by

$$\chi_L(\lambda) \stackrel{\text{def}}{=} \sum_{x \in L} \mu(\hat{0}, x)\, \lambda^{\rho(\hat{1}) - \rho(x)}.$$

The form of $A(x)$ in Example 2a shows that $S(\hat{0})$ is, up to a factor of a power of q, equal to $\chi_L(q^k)$. Now Zaslavsky [25, §2] has shown that if L is the intersection lattice of a real hyperplane arrangement $\mathscr{H}$, then $|\chi_L(-1)|$ and $|\chi_L(1)|$ give the total number of regions and the total number of bounded regions respectively. Moreover, Athanasiadis [1] has shown that if all the hyperplanes are defined over the rationals, then it suffices to consider the "localizations" of $\mathscr{H}$ and χ_L to a finite field over a sufficiently large prime q. In short, estimating $S(\hat{0})$ allows us to count regions in hyperplane arrangements.

EXAMPLE 2b. Let Γ be a finite undirected graph with vertex set V, and let n be a positive integer. Let $\mathscr{A}$ be the set of all *colorings of V with at most n colors*, i.e., the set of all maps $\kappa: V \to \{1, 2, ..., n\}$. Let L *be the lattice of contractions* of Γ, i.e., the collection of all set partitions π of V such that each block of π induces a *connected* subgraph of Γ. The partial order of L is given by reverse refinement, so that $\hat{0}$ is the partition where each block is a singleton. Let $\sigma(\kappa)$ be the maximal element $\pi \in L$ such that κ colors each block of π monochromatically. It is not hard to show that

$$A(\pi) = n^{|\pi|},$$

where $|\pi|$ denotes the number of blocks of π, and that $S(\hat{0})$ is the *chromatic polynomial* of Γ, i.e., the number of colorings of V with at most n colors such that adjacent vertices are always assigned different colors.

It is not immediately obvious that Example 2b is really a special case of Example 2, but it is. In fact, it is a special case of Example 2a. This is explained in [3]; alternatively, for readers familiar with matroids, it follows because every graphic matroid is representable over every field [16, Proposition 5.1.2].

A number of important problems in combinatorics involve computing chromatic polynomials. For instance, if $G = K_n \times K_n$, the graph whose vertex set V consists of the points of an $n \times n$ grid and in which two vertices are adjacent if and only if they lie in the same row or column, then $S(\hat{0})$ is just the number of $n \times n$ Latin squares. (A Latin square is an $n \times n$ array

of integers such that each row and each column is a permutation of the integers from 1 to n.) Computing the asymptotic number of $n \times n$ Latin squares is a major open problem. The best partial result is due to Godsil and McKay [7], who obtain an asymptotic formula for the number of $k \times n$ Latin rectangles when $k = o(n^{6/7})$.

EXAMPLE 3. Let G be an $n \times n$ grid and let B be a subset of G. A *rook placement* R on B is a subset of B such that no two elements of R lie in the same row or column of B. (The terminology comes from the problem of putting rooks on a chessboard such that no two rooks can take each other.) Let $\mathscr{A}$ be the set of all rook placements $R \subseteq G$ on G such that $|R| = n$. Clearly $|\mathscr{A}| = n!$ because such rook placements are in bijection with permutations of n. Let L be the set of all rook placements on B (partially ordered by inclusion) together with an adjoined maximum element $\hat{1}$. Define σ by setting $\sigma(R) = R \cap B$. Then $A(R) = (n - |R|)!$ and $S(\hat{0})$ is the number of rook placements $R \subseteq G$ on G such that $|R| = n$ and no element of R lies in B.

Classical accounts of rook theory may be found in [23, Chapter 2] and [20, Chapters 7–8]. Although rook theory is an old branch of combinatorics, there have been a number of recent advances that have rejuvenated it, e.g., [2, 5, 6, 8–10]. There also exist q-analogs of rook theory, e.g., [4], that fit nicely into our framework, but we shall omit the details.

In number theory, the lattice L is always a Boolean algebra, and hence number theorists tend to concentrate on the structure of $\mathscr{A}$ rather than the structure of L. In contrast, combinatorialists deal with a wide variety of lattices and tend to focus on the structure of L, introducing $\mathscr{A}$ and σ only when they are needed in the course of a proof. Our definition of a sieve puts $\mathscr{A}$ and L on an equal footing.

3. SELBERG'S Λ^2 UPPER BOUND METHOD

The notion of a sieve is very general, and further conditions are needed to produce nontrivial results.

DEFINITION. Let L be a lattice with a minimum element $\hat{0}$. A function $f: L \to \mathbb{R}\backslash\{0\}$ is *multiplicative* if $f(\hat{0}) = 1$ and $f(x \vee y)\, f(x \wedge y) = f(x)\, f(y)$ for all $x, y \in L$.

Most number-theoretic problems satisfy the following condition.

The Multiplicativity Condition. There exists a multiplicative function f and a "small" remainder function R: $L \to \mathbb{R}$ such that

$$A(x) = \frac{A(\hat{0})}{f(x)} + R(x) \tag{3.1}$$

for all $x \in L$.

The word "small" in placed in scare quotes because its meaning varies somewhat from problem to problem. In general, the smaller R is the better the results, and if R exceeds a certain size threshold then it swamps the main term and no results can be obtained.

The multiplicativity condition holds in Example 1 above, as we may see as follows. Let $N(x)$ denote the number of solutions of

$$h(v) = 0 \pmod{x}$$

in the range $1 \leqslant v \leqslant x$. By the Chinese Remainder Theorem, $N(x)$ is multiplicative. If we set $f(x) = x/N(x)$ then it is not hard to show that

$$A(x) = \frac{A(\hat{0})}{f(x)} + R(x)$$

with

$$|R(x)| \leqslant \frac{x}{f(x)}.$$

That this is indeed "small" may be seen heuristically by noting that $|R(x)|$ is roughly constant whereas $A(x)$ grows as n grows.

Now let us return to the general case. Write $\mu(x)$ for $\mu(\hat{0}, x)$ for simplicity. If (3.1) holds, then

$$S(\hat{0}) = A(\hat{0}) \sum_{x \in L} \frac{\mu(x)}{f(x)} + \sum_{x \in L} \mu(x)\, R(x). \tag{3.2}$$

Now, it is natural to think of the first sum in (3.2) as the main term and the second sum in (3.2) as the error term, since $R(x)$ is small. It is also natural to try to estimate both the main term and the error term directly. But typically, the problem with doing this is that even though $R(x)$ is "small," there are a lot of terms in the summation, making the error term unacceptably large.

Brun's key idea was to perturb the Möbius function by setting it equal to zero for selected arguments. Done carefully, this drastically reduces the number of summands in the error term without disturbing the main term

too much. Selberg later improved Brun's results by allowing himself the flexibility of perturbing the Möbius function in more general ways.

More precisely, following Selberg [22], define a *Λ-system* to be a function $\lambda: L \to \mathbb{R}$. If

$$\sum_{x \preccurlyeq y} \lambda(x) \leqslant \sum_{x \preccurlyeq y} \mu(x) \tag{3.3}$$

for all $y \in L$, then we say that λ is a *lower bound Λ-system*. Similarly, if

$$\sum_{x \preccurlyeq y} \mu(x) \leqslant \sum_{x \preccurlyeq y} \lambda(x) \tag{3.4}$$

for all $y \in L$, then we say that λ is an *upper bound Λ-system*. The reason for this nomenclature is the following.

PROPOSITION 2. *Let $(\mathscr{A}, w, L, \mu)$ be a weighted sieve. Assume that* (3.1) *holds. Let λ^- be a lower bound Λ-system and let λ^+ be an upper bound Λ-system. Then*

$$A(\hat{0}) \sum_{x \in L} \frac{\lambda^-(x)}{f(x)} + \sum_{x \in L} \lambda^-(x)\, R(x) \leqslant S(\hat{0}) \leqslant A(\hat{0}) \sum_{x \in L} \frac{\lambda^+(x)}{f(x)} + \sum_{x \in L} \lambda^+(x)\, R(x).$$

Proof. From the proof of Proposition 1 we have

$$S(\hat{0}) = \sum_{a \in \mathscr{A}} w(a) \sum_{x \preccurlyeq \sigma(a)} \mu(x),$$

so by (3.3) and (3.4),

$$\sum_{a \in \mathscr{A}} w(a) \sum_{x \preccurlyeq \sigma(a)} \lambda^-(x) \leqslant S(\hat{0}) \leqslant \sum_{a \in \mathscr{A}} w(a) \sum_{x \preccurlyeq \sigma(a)} \lambda^+(x).$$

Thus

$$\sum_{x \in L} \lambda^-(x) \sum_{\{a \in \mathscr{A} : x \preccurlyeq \sigma(a)\}} w(a) \leqslant S(\hat{0}) \leqslant \sum_{x \in L} \lambda^+(x) \sum_{\{a \in \mathscr{A} : x \preccurlyeq \sigma(a)\}} w(a)$$

or

$$\sum_{x \in L} \lambda^-(x)\, A(x) \leqslant S(\hat{0}) \leqslant \sum_{x \in L} \lambda^+(x)\, A(x).$$

The conclusion now follows from (3.1). ∎

The idea now is to find Λ-systems λ for which $\lambda(x)$ is zero for many values of x (so that the error term is cut down to a manageable size) yet for which the inequalities in Proposition 2 are not too loose and for which the main term can still be estimated accurately.

Finding good Λ-systems can involve a lot of technical complications and this is not the place to delve deeply into all the variations that number theorists have developed (see [12] or [22] for full accounts, and [13, Chapter IV] or [17] for introductions). However, to justify our claim that much number theory carries over to our generalized setting, we present here Selberg's Λ^2 upper bound method. This is perhaps the result that involves the fewest technicalities, yet it is still very powerful.

THEOREM 1. *Let $(\mathscr{A}, w, L, \sigma)$ be a weighted sieve, and let X be a non-empty order ideal of L. Assume that the multiplicativity condition holds. Let $g: L \to \mathbb{R}$ be defined by*

$$g(y) = \sum_{x \leqslant y} \mu(x, y)\, f(x).$$

Assume that $g(y) \neq 0$ for all y, and let

$$Q = \sum_{x \in X} \frac{\mu^2(x)}{g(x)}.$$

Then

$$S(\hat{0}) \leqslant \frac{A(\hat{0})}{Q} + \sum_{x_1, x_2 \in X} \ell(x_1)\, \ell(x_2)\, R(x_1 \vee x_2),$$

where

$$\ell(x) = \frac{f(x)}{Q} \sum_{\{y \in X : x \leqslant y\}} \frac{\mu(x, y)\, \mu(y)}{g(y)}$$

when $x \in X$ and $\ell(x) = 0$ if $x \notin X$.

Proof. We remark first of all that if $Q = 0$ then the division by Q should be interpreted as being "infinity" so that the theorem is vacuous.

The proof carries over almost word for word from the number-theoretic case. Define

$$X^* \stackrel{\text{def}}{=} \{x: x = y \vee z \text{ for some } y, z \in X\}.$$

Now let $f: X \to \mathbb{R}$ be *any* map such that $\ell(\hat{0}) = 1$. We claim that the Λ-system λ_ℓ defined by

$$\lambda_\ell(x) \stackrel{\text{def}}{=} \sum_{\substack{x_1, x_2 \in X \\ x_1 \vee x_2 = x}} \ell(x_1)\, \ell(x_2)$$

for $x \in X^*$ and $\lambda_\ell(x) = 0$ otherwise is necessarily an upper bound Λ-system. For

$$\sum_{x \leqslant y} \lambda_\ell(x) = \sum_{x \leqslant y} \sum_{\substack{x_1, x_2 \in X \\ x_1 \vee x_2 = x}} \ell(x_1)\, \ell(x_2) = \left(\sum_{z \in X : z \leqslant y} \ell(z) \right)^2,$$

and this is nonnegative for all y and equal to one when $y = \hat{0}$ (because $\ell(\hat{0}) = 1$), so (3.4) holds.

Define H_ℓ by

$$H_\ell \stackrel{\text{def}}{=} \sum_{x \in X^*} \frac{\lambda_\ell(x)}{f(x)}.$$

By Möbius inversion,

$$f(x) = \sum_{y \leqslant x} g(y),$$

so by the multiplicativity of f,

$$\frac{1}{f(x_1 \vee x_2)} = \frac{1}{f(x_1)\, f(x_2)} \sum_{y \leqslant x_1 \wedge x_2} g(y).$$

Hence

$$\begin{aligned} H_\ell &= \sum_{x_1 \in X} \sum_{x_2 \in X} \frac{\ell(x_1)\, \ell(x_2)}{f(x_1 \vee x_2)} \\ &= \sum_{x_1 \in X} \sum_{x_2 \in X} \frac{\ell(x_1)\, \ell(x_2)}{f(x_1)\, f(x_2)} \sum_{y \leqslant x_1 \wedge x_2} g(y) \\ &= \sum_{y \in X} g(y) \left(\sum_{x \in X : y \leqslant x} \frac{\ell(x)}{f(x)} \right)^2. \end{aligned}$$

Now define

$$k(y) \stackrel{\text{def}}{=} \sum_{x \in X : y \leqslant x} \frac{\ell(x)}{f(x)}.$$

By Möbius inversion on X (which has the same Möbius function as L because X is an order ideal),

$$\ell(x) = f(x) \sum_{y \in X : x \leqslant y} \mu(x, y)\, k(y).$$

Since f is multiplicative, $f(\hat{0}) = 1$, so the condition $\ell(\hat{0}) = 1$ implies

$$\sum_{y \in X} \mu(y)\, k(y) = 1.$$

We therefore have the identity

$$H_\ell = \sum_{y \in X} g(y)\, k(y)^2 = \sum_{y \in X} \frac{1}{g(y)} \left(g(y)\, k(y) - \frac{\mu(y)}{Q} \right)^2 + \frac{1}{Q}.$$

Thus if we take ℓ to be as in the statement of the theorem, then $\ell(\hat{0}) = 1$, and

$$k(y) = \frac{1}{Q} \frac{\mu(y)}{g(y)},$$

so $H_\ell = 1/Q$. The theorem then follows from Proposition 2, since λ_ℓ is an upper bound Λ-system. ∎

The smaller X is, the fewer the summands in the error term but the poorer the approximation. The upper bound in Theorem 1 is not the best of all possible upper bounds, but it is simple and explicit and in practice a judicious choice of X leads to quite a sharp approximation, at least in many number-theoretic applications.

4. APPLICATIONS TO COMBINATORICS

The combinatorial examples described previously are all potential applications for Theorem 1 (or similar theorems borrowed from number theory). However, we should point out one serious difficulty. In order to apply Theorem 1, we must posit the multiplicativity condition. But for many of the lattices that appear in combinatorics, there are *no* nonconstant multiplicative functions. For example, it is easy to check that the lattice of contractions of any graph that contains a cycle of length four or more admits no nonconstant multiplicative functions.

Complete despair at this obstacle would be premature, however. By taking logarithms, we see that the notion of a multiplicative function is equivalent to the notion of a *valuation*, i.e., a function $v: L \to \mathbb{R}$ such that

$$v(x \vee y) + v(x \wedge y) = v(x) + v(y).$$

Valuations do exist on many important lattices and they have been studied by a number of people.

Perhaps more importantly, if we carefully re-examine the discussion in the previous section, we see that the crucial notion of a Λ-system does *not* depend on the multiplicativity of f but only on the algebraic form of (3.1). Therefore we can still play the game of seeking clever upper and lower bound Λ-systems even if no multiplicative functions exist. Now, without *some* condition on f, we cannot expect to obtain nontrivial bounds. However, note that most of the lattices that arise in combinatorics are at least *semimodular*, i.e., they are graded and the rank function satisfies

$$\rho(x \vee y) + \rho(x \wedge y) \leqslant \rho(x) + \rho(y).$$

In semimodular lattices there will always be "submultiplicative" functions, and these might be reasonably good substitutes for multiplicative functions.

The logical next step in the program initiated here would be to analyze particular applications in detail. We hope to carry this out in future papers.

As a final remark, we mention that sieve methods in number theory often need to be supplemented by analysis. Theorems and conjectures about zeta functions and L-functions, for example, enter naturally into many problems. It is not clear what would play this role in combinatorics, but a natural candidate would be the Rédei zeta function [15] or possibly the zeta function of an arithmetical semigroup [14]. The suggestion in [15] that the analytical properties of the Rédei zeta function be worked out does not seem to have been pursued yet; it would be interesting to do this and to examine the connection with sieve methods.

ACKNOWLEDGMENTS

Anyone who has read Gian-Carlo Rota's papers will recognize the profound influence that his ideas have had on this paper. I am also grateful for his personal encouragement and enthusiasm, without which the ideas in this paper would probably have remained forever unwritten and disorganized in my brain. I thank Trevor Wooley for pointing me in the right direction with regard to the number-theoretic literature, and Greg Martin for helpful discussions. This work was supported in part by a National Science Foundation postdoctoral fellowship.

Note added in proof. I just learned of a paper by R. J. Wilson, The Selberg sieve for a lattice, *in* "Combinatorial Theory and Its Applications, Balatonfüred (Hungary)," *Colloq. Math. Soc. János Bolyai* **4** (1969), 1141–1149. Wilson anticipates some of the ideas in this paper but treats them from a slightly different point of view.

REFERENCES

1. C. Athanasiadis, Characteristic polynomials of subspace arrangements and finite fields, *Adv. Math.* **122** (1996), 193–233.
2. A. Björner, L. Lovász, S. T. Vrećica, and R. T. Živaljević Chessboard complexes and matching complexes, *J. London Math. Soc.* **49** (1994), 25–39.
3. A. Brini, Some remarks on the critical problem, *in* "Matroid Theory and Its Applications: III ciclo 1980, Villa Monastero, Varenna-Como" (A. Barlotti, Ed.), Napoli, 1982.
4. W. Y. C. Chen and G.-C. Rota, q-analogs of the inclusion-exclusion principle and permutations with restricted position, *Discrete Math.* **104** (1992), 7–22.
5. K. Ding and P. Terwilliger, On Garsia-Remmel problem of rook equivalence, *Discrete Math.* **149** (1996), 59–68.
6. M. Dworkin, Factorization of the cover polynomial, *J. Combin. Theory Ser. B* **71** (1997), 17–53.
7. C. D. Godsil and B. D. McKay, Asymptotic enumeration of Latin rectangles, *J. Combin. Theory Ser. B* **48** (1990), 19–44.
8. J. Haglund, Rook theory and hypergeometric series, *Adv. Appl. Math.* **17** (1996), 408–459.
9. J. Haglund, Rook theory, compositions, and zeta functions, *Trans. Amer. Math. Soc.* **348** (1996), 3799–3825.
10. J. Haglund, K. Ono, and L. Sze, Rook theory and t-cores, *J. Combin. Theory Ser. A*, to appear.
11. G. Hajós, Über einfache und mehrfache Bedeckung des n-dimensionalen Raumes mit einem Wurfelgitter, *Math. Z.* **47** (1942), 427–467.
12. H. Halberstam and H.-E. Richert, "Sieve Methods," Academic Press, New York, 1947.
13. H. Halberstam and K. F. Roth, "Sequences," Springer-Verlag, New York, 1983.
14. J. Knopfmacher, "Abstract Analytic Number Theory," Dover, New York, 1990.
15. J. P. S. Kung, M. R. Murty, and G.-C. Rota, On the Rédei zeta function, *J. Number Theory* **12** (1980), 421–436.
16. J. G. Oxley, "Matroid Theory," Oxford Univ. Press, New York, 1992.
17. C. Pomerance and A. Sárközy, Combinatorial number theory, *in* "Handbook of Combinatorics" (R. Graham, M. Grötschel, and L. Lovász, Eds.), Chap. 20, Elsevier, New York, 1995.
18. L. Rédei, Zetafunktionen in der Algebra, *Acta Math. Acad. Sci. Hungar.* **6** (1955), 5–25.
19. L. Rédei, Die gruppentheoretischen Zetafunktionen und der Satz von Hajós, *Acta Math. Acad. Sci. Hungar.* **6** (1955), 271–279.
20. J. Riordan, "An Introduction to Combinatorial Analysis," Wiley, New York, 1958.
21. G.-C. Rota, On the foundations of combinatorial theory. I. Theory of Möbius functions, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **2** (1964), 340–368.
22. A. Selberg, Lectures on sieves, *in* "Collected Papers" (A. Selberg, Ed.), Vol. II, Chap. 45, Springer-Verlag, New York, 1991.
23. R. P. Stanley, "Enumerative Combinatorics," Vol. 1, Cambridge Stud. Adv. Math. 49, Cambridge Univ. Press, New York, 1997.
24. D. G. Wagner, The Tutte dichromate and Whitney homology of matroids, preprint.
25. T. Zaslavsky, Facing up to arrangements: Face-count formulas for partitions of space by hyperplanes, *Mem. Amer. Math. Soc.* **1**, No. 154 (1975).