STERN NOTES, MATH 595, SPRING 2012

BRUCE REZNICK, UIUC

3. Linear recurrences

3.1. **Basics.** In this section, we talk about constant-coefficient linear recurrences; all objects should be viewed as living in $\mathbb{C}$. Recurrences usually arise in one of two forms. The first is the direct one: suppose that $(a(n))$ is a sequence which satisfies the *homogeneous d-th order linear recurrence*:

$$a(n) + \sum_{j=1}^{d} c_j a(n-j) = 0, \qquad n \geq d, \tag{3.1}$$

or, equivalently,

$$a(n+d) = -\sum_{j=1}^{d} c_j a(n+d-j) = 0, \qquad n \geq 0. \tag{3.2}$$

We say that the undetermined values of $a(i)$, $0 \leq i \leq d-1$, are the *initial conditions*. Associated to the recurrence (3.1) is the *characteristic polynomial*

$$\phi(z) = z^d + \sum_{j=1}^{d} c_j z^{d-j}. \tag{3.3}$$

If $\phi(t_0) = 0$, then $a(n) = t_0^n$ is easily seen to satisfy (3.1). We also make the fairly obvious point that if (3.1) holds for $(a(n))$, then so does

$$a(n+k) + \sum_{j=1}^{d} c_j a(n+k-j) = 0, \qquad n \geq d, \tag{3.4}$$

for $k \geq 1$, and by adding together identities such as (3.4), it is not hard to prove that, if $\zeta(z) = \phi(z)\eta(z)$, then $(a(n))$ also satisfies the recurrence whose characteristic polynomial is ζ.

The second way recurrences arise is as a *matrix system*: d sequences $(a_j(n))$, $1 \le j \le d$, which are related by:

$$\begin{aligned} a_j(n+1) &= \sum_{k=1}^{d} m_{jk} a_k(n), \qquad n \ge 0, \quad 1 \le j \le d, \\ \begin{pmatrix} a_1(n+1) \\ \cdots \\ a_d(n+1) \end{pmatrix} &= \begin{pmatrix} m_{11} & \cdots & m_{1d} \\ \cdots & \cdots & \cdots \\ m_{d1} & \cdots & m_{dd} \end{pmatrix} \begin{pmatrix} a_1(n) \\ \cdots \\ a_d(n) \end{pmatrix}. \end{aligned} \tag{3.5}$$

More formally, let $A(n) = (a_1(n) \cdots a_d(n))^T$ be the column vector of sequences and let $M = [m_{jk}]$ be the matrix of coefficients. Then (3.5) becomes

$$A(n+1) = MA(n) \implies A(n) = M^n A(0), \quad n \ge 0. \tag{3.6}$$

Here, $A(0)$ provides the initial condition.

Although the methods of solutions for these two kinds of recurrences are different, they can each be transformed to the other. The Cayley-Hamilton Theorem states that if $\phi(\lambda) = \det(\lambda I_d - M)$ is the characteristic polynomial of M, then $\phi(M) = 0$, where "0" is construed as the $d \times d$ matrix of 0's. Supposing ϕ is given by (3.3) for convenience, we then have

$$\begin{aligned} \phi(t) = t^d + c_1 t^{d-1} + \cdots + c_d &\implies M^d + c_1 M^{d-1} + \cdots + c_d I_d = 0 \\ &\implies M^{n+d} + c_1 M^{n+d-1} + \cdots + c_d M^n = 0 \\ &\implies A(n+d) + c_1 A(n+d-1) + \cdots + c_d A(d) = 0, \end{aligned} \tag{3.7}$$

where the final "0" is the zero column vector. The last equation in (3.7) is simply the assertion that each $a_j(n)$ satisfies (3.1).

On the other hand, (3.2) can be simple-mindedly rewritten as:

$$\begin{pmatrix} a(n+1) \\ a(n+2) \\ \cdots \\ a(n+d-1) \\ a(n+d) \end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 1 \\ -c_d & -c_{d-1} & \cdots & -c_2 & -c_1 \end{pmatrix} \begin{pmatrix} a(n) \\ a(n+1) \\ \cdots \\ a(n+d-2) \\ a(n+d-1) \end{pmatrix}. \tag{3.8}$$

The matrix in (3.8) (or its transpose) is sometimes called the *companion matrix* to the polynomial ϕ, and has characteristic polynomial $(-1)^d \phi$.

3.2. **Solving recurrences.** For completeness' sake, we include a self-contained proof of the method of Partial Fractions, which requires the Fundamental Theorem of Algebra to factor the denominator, but no explicit theory of complex variables.

Suppose

$$F(z) = \frac{p(z)}{q(z)} = \frac{b_n z^n + b_{n-1} z^{n-1} + \cdots + b_0}{c_m z^m + c_{m-1} z^{m-1} + \cdots + c_0} \tag{3.9}$$

is a rational function, where $b_k, c_k \in \mathbb{C}$, $b_n c_m \neq 0$ and $m > n$. Assume that F is presented in lowest terms and $c_m = 1$. Suppose further that

$$q(z) = \prod_{j=1}^{r} (z - z_j)^{m_j}, \tag{3.10}$$

where $z_j \in \mathbb{C}$, the z_j's are distinct, $m_j \in \mathbb{N}$ with $\sum m_j = m$, and $p(z_j) \neq 0$. Then as we tell our calculus students:

Theorem 3.1. *Suppose F is given by* (3.9)*, and* (3.10) *holds. Then there exist $r_{j\ell} \in \mathbb{C}$ so that*

$$F(z) = \sum_{j=1}^{r} \left(\frac{r_{j1}}{z - z_j} + \cdots + \frac{r_{jm_j}}{(z - z_j)^{m_j}} \right). \tag{3.11}$$

Conversely, if F is given by (3.11)*, then F is a rational function of shape* (3.9) *for which $n < m$, and* (3.10) *hold.*

Proof. We induct on $m = \deg(q)$. If $m = 1$, then $n = 0$ and there is nothing to prove. Suppose the theorem is valid for q with $\deg q \leq m - 1$ and suppose $q(z) = (z - z_1)^{m_1} \bar{q}(z)$, with $\bar{q}(z_1) \neq 0$. We do not rule out the possibility that $\bar{q}(z) \equiv 1$. Consider the expression

$$F(z) - \frac{\alpha}{(z - z_1)^{m_1}} = \frac{p(z) - \alpha \bar{q}(z)}{q(z)}.$$

Let $p_\alpha(z) = p(z) - \alpha \bar{q}(z)$; if $\alpha_0 = \frac{p(z_1)}{\bar{q}(z_1)}$, then $p_{\alpha_0}(z_1) = 0$, hence $p_{\alpha_0}(z) = (z - z_1)\hat{p}(z)$. It follows that

$$F(z) = \frac{\alpha_0}{(z - z_1)^{m_1}} + \frac{\hat{p}(z)}{(z - z_1)^{m_1 - 1} \bar{q}(z)},$$

and a partial fraction expression for $F(z) - \frac{\alpha_0}{(z - z_1)^{m_1}}$ exists by the inductive hypthesis.

For the converse, if F is given by (3.11), then multiplication by $q(z)$ yields a polynomial on the right-hand side, with degree at most $m - 1$. □

We now return to (3.1) and add a subtle new hypothesis:

$$c_d \neq 0. \tag{3.12}$$

(This retriction actually offers no practical limitations. Suppose (3.12) is not satisfied. If $c_j = 0$ for all j, then (3.1) has only the zero solution. Otherwise, suppose $c_e \neq 0$ and $c_j = 0$ for $e + 1 \leq j \leq d$. Let $\tilde{a}(n) = a(n + d - e), n \geq 0$. Then (3.1) becomes

$$\tilde{a}(n) + \sum_{j=1}^{e} c_j \tilde{a}(n - j) = 0, \qquad n \geq e,$$

with no constraint involving $a(k)$ for $0 \leq k < d - e$. In this case, first solve for $\tilde{a}$ by the algorithm described below, and then write $a(n) = \tilde{a}(n - (d - e))$ for $n \geq d - e$.)

It is now convenient to define

$$\psi(z) = 1 + \sum_{j=1}^{d} c_j z^j = z^d \phi(z^{-1}),$$

as a polynomial with true degree d (by (3.12)) and note that $z_j \neq 0$ in (3.10) and

$$\phi(z) = \prod_{j=1}^{r} (z - z_j)^{m_j} \iff \psi(z) = \prod_{j=1}^{r} (1 - z z_j)^{m_j}.$$

Let $M = \max(1, \sum_j |c_j|) \geq 1$. If $|a(i)| \leq T$ for $i = 0, \ldots, d-1$, then

$$|a(d)| \leq \sum_{j=1}^{d} |c_j||a(d-j)| \leq T \sum_{j=1}^{d} |c_j| \leq TM,$$

and so $|a(i)| \leq MT$ for $i = 1, \ldots, d$. An easy induction implies that $|a(n)| \leq M^{n+1-d}T$ for each $n \geq d$, and so the generating function for $(a(n))$ will be an analytic function with radius of convergence $\geq M^{-1}$. Let

$$f(z) := \sum_{n=0}^{\infty} a(n) z^n.$$

Then, as we saw in (2.4),

$$\psi(z) f(z) = \sum_{n=0}^{d-1} \left(a(n) + \sum_{j=1}^{n} c_j a(n-j) \right) z^n := p(z), \tag{3.13}$$

so $f(z)$ is a rational function:

$$f(z) = \frac{p(z)}{\psi(z)} = \frac{p(z)}{\prod_{j=1}^{r} (1 - z_j z)^{m_j}}, \tag{3.14}$$

where $\deg(p) \leq d - 1 < d$.

It follows from Theorem 3.1 that there exist $r_{j\ell} \in \mathbb{C}$ such that

$$\sum_{n=0}^{\infty} a_n z^n = \sum_{j=1}^{r} \left(\frac{r_{j1}}{1 - z_j z} + \cdots + \frac{r_{jm_j}}{(1 - z_j z)^{m_j}} \right). \tag{3.15}$$

The power series for the right-hand side was already computed in (2.18):

$$\frac{1}{1 - \lambda z} = \sum_{n=0}^{\infty} \lambda^n z^n, \quad \frac{1}{(1 - \lambda z)^{r+1}} = \sum_{n=0}^{\infty} \binom{n+r}{r} \lambda^n z^n, \ r \geq 1.$$

Thus, the coefficient of z^n in the j-th summand in (3.15) can be expressed as follows:

$$\left(r_{j1} + \sum_{\ell=2}^{m_j} r_{j\ell} \cdot \frac{(n+1) \cdots (n+\ell-1)}{(\ell-1)!} \right) z_j^n = p_j(n) z_j^n, \tag{3.16}$$

where p_j is a polynomial with degree $\leq m_j - 1$. The coefficients of p_j depend on the r_ℓ's, which depend on the initial conditions of the recurrence. We have therefore proved the main theorem about linear recurrences.

Theorem 3.2. *If $(a(n))$ is a sequence satisfying (3.1) with $c_d \neq 0$, and if*

$$\phi(z) = z^d + \sum_{j=1}^{k} c_j z^{d-j} = \prod_{j=1}^{r} (z - z_j)^{m_j}, \tag{3.17}$$

then there exist polynomials p_j so that for $n \geq 0$,

$$a(n) = \sum_{j=1}^{r} p_j(n) z_j^n, \quad \deg(p_j) \leq m_j - 1. \tag{3.18}$$

Conversely, any sequence $(a(n))$ defined by (3.18) satisfies the recurrence (3.1).

The proof of the last assertion is that (3.18) implies that f is given by (3.15), and so ψf is a polynomial. In practice, "most" polynomials have distinct roots, so the polynomials p_j are, in fact, "usually" constants.

As noted in Chapter 2, the set of sequences satisfying a recurrence such as (3.1) forms a d-dimensional vector space. One natural basis follows from (3.18), namely

$$\left\{ n^i z_j^n : \quad 1 \leq j \leq r,\ 0 \leq i \leq m_j - 1 \right\}.$$

A somewhat more natural basis comes from (3.14) by considering those sequences whose generating functions are given by $\frac{z^i}{\psi(z)}$ for $0 \leq i \leq d-1$. Define $b_0(n)$ to be the sequence satisfying (3.1) with initial conditions

$$b_0(0) = \cdots = b_0(d-2) = 0, \quad b_0(d-1) = 1.$$

It follows from (3.13) that

$$\sum_{n=0}^{\infty} b_0(n) z^n = \frac{z^{d-1}}{\psi(z)}.$$

Upon dividing (3.2) by z^k, $1 \leq k \leq d-1$, we see that

$$\sum_{n=0}^{\infty} b_0(n+k) z^n = \frac{z^{d-1-k}}{\psi(z)}.$$

In other words, the sequences $\{(b_0(n)), (b_0(n+1)), \ldots, (b_0(n+d-1))\}$ form a basis for this subspace. This is familiar in the Fibonacci sequence setting.

There isn't too much to say about solving (3.6) directly. The standard methodology is to put the matrix in Jordan canonical form:

$$M = C^{-1} D C \implies M^n = C^{-1} D^n C. \tag{3.19}$$

If the characteristic polynomial of M has distinct roots, then D is diagonal and D^n is easy to calculate; otherwise, D is a block diagonal matrix.

3.3. **Standard examples.** Here is a simple example of a recurrence with a repeated root. Let

$$a(n) = 4a(n-1) - 4a(n-2); \quad a(0) = r,\ a(1) = 2s; \quad \phi(z) = (z-2)^2. \tag{3.20}$$

Then $a(2) = 4(2s-r), a(3) = 4(8s-4r) - 4(2s) = 24s - 16r = 8(3s-2r)$, and

$$\begin{aligned} (1-4z+4z^2)\sum_{n=0}^{\infty} a(n)z^n &= a(0) + (a(1) - 4a(0))z \\ +\sum_{n=2}^{\infty}(a(n) - 4a(n-1) + 4a(n-2))z^n &= r + (2s-4r)z. \end{aligned} \tag{3.21}$$

Thus,

$$\begin{aligned} \sum_{n=0}^{\infty} a(n)z^n &= \frac{r + (2s-4r)z}{(1-2z)^2} = \frac{2r-s}{1-2z} + \frac{-r+s}{(1-2z)^2} \\ \implies a(n) &= ((2r-s) + (s-r)(n+1))\,2^n = 2^n(ns - (n-1)r), \end{aligned} \tag{3.22}$$

as can verified by the first few terms of the series given above.

Now the inevitable Fibonacci example. The Fibonacci numbers are defined by $F_0 = 0, F_1 = 1, F_n = F_{n-1} + F_{n-2}$, or $F_n - F_{n-1} - F_{n-2} = 0$ for $n \geq 2$. Following the previous procedure, we see that

$$\sum_{n=0}^{\infty} F_n z^n = \frac{F_0 + (F_1 - F_0)z}{1 - z - z^2} = \frac{z}{1-z-z^2}.$$

Since $z^2 - z - 1 = (z-\phi)(z-\bar{\phi})$, where $\phi = \frac{1+\sqrt{5}}{2} \approx 1.618$ and $\bar{\phi} = \frac{1-\sqrt{5}}{2} \approx -.618$, there exist constants c_j so that $F_n = c_1\phi^n + c_2\bar{\phi}^n$. The initial conditions imply $c_1 + c_2 = 0$ and $c_1\phi + c_2\bar{\phi} = 1$, yielding the *Binet formula*:

$$F_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right). \tag{3.23}$$

Since $|\bar{\phi}| < 1$, we see that $F_n \approx \frac{\phi^n}{\sqrt{5}}$; in fact, F_n is the closest integer to $\frac{\phi^n}{\sqrt{5}}$ for $n \geq 0$. It is also easy to extend the definition of the Fibonacci sequence to negative n; the equation $\phi\bar{\phi} = -1$ is instrumental in showing that $F_{-n} = (-1)^{n-1}F_n$.

It follows from the geometric series that

$$\begin{aligned} \sum_{n=0}^{\infty} F_n z^n = \frac{z}{1-z-z^2} = z\sum_{m=0}^{\infty}(z+z^2)^m &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{i+j}{j} z\cdot z^i(z^2)^j \\ \implies F_n &= \sum_{j\geq 0}\binom{n-1-j}{j}. \end{aligned} \tag{3.24}$$

(The binomial coefficient shuts off the final sum at $j = \lfloor \frac{n-1}{2} \rfloor$.) This formula lets you find the Fibonacci numbers by summing along a slope of Pascal's triangle, and will show up in the next chapter.

Closely related are the *Lucas* numbers, defined by $L_0 = 2, L_1 = 1$, and $L_n = L_{n-1} + L_{n-2}$ for $n \geq 2$. Since (L_n), (F_n) and (F_{n+1}) are three sequences with the same second-order recurrence, they are linearly dependent. Thus, there exist constants c_i so that $c_1 L_n + c_2 F_n + c_3 F_{n+1} = 0$, and (F_n) and (F_{n+1}) are not proportional, so we may take $c_1 = 1$. Putting $n = 0, 1$, we see that

$$\begin{aligned} 2 + c_3 = 1 + c_2 + c_3 = 0 &\implies c_2 = 1, c_3 = -2 \\ &\implies L_n = 2F_{n+1} - F_n = F_{n-1} + F_{n+1}. \end{aligned}$$

Similarly, (F_n), (L_n) and (L_{n+1}) are linearly dependent, and $F_n = \frac{1}{5}(2L_{n+1} - L_n)$. The similarity of coefficients is not accidental; it's an exercise that for fixed α, β, m,

$$L_{n+m} = \alpha F_{n+1} + \beta F_n \iff F_{n+m} = \frac{1}{5}\left(\alpha L_{n+1} + \beta L_n\right).$$

For a fixed positive integer m, there must be a dependence among (F_n), (F_{n+1}) and (F_{n+m}). Taking $n = 0, 1$ in the equation $F_{n+m} = \alpha F_n + \beta F_{n+1}$ implies that $F_m = \beta, F_{m+1} = \alpha + \beta$, so it's easy to derive the *Fibonacci addition formula*:

$$F_{n+m} = (F_{m+1} - F_m)F_n + F_m F_{n+1} = F_{m+1}F_n + F_m F_{n-1}. \tag{3.25}$$

It is also easy to show that $L_n = \phi^n + \bar{\phi}^n$ and $F_n L_n = F_{2n}$. Finally, observe that

$$\begin{aligned} (x+y)^n + (x-y)^n &= 2\sum_{k\geq 0} \binom{n}{2k} x^{n-2k} y^{2k}; \\ (x+y)^n - (x-y)^n &= 2\sum_{k\geq 0} \binom{n}{2k+1} x^{n-2k-1} y^{2k+1}. \end{aligned} \tag{3.26}$$

Taking $x = \frac{1}{2}$ and $y = \frac{\sqrt{5}}{2}$, we find that $x + y = \phi$ and $x - y = \bar{\phi}$, and in view of the formulas for F_n and L_n, it follows from the binomial theorem that

$$L_n = \frac{1}{2^{n-1}} \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} 5^k, \qquad F_n = \frac{1}{2^{n-1}} \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \binom{n}{2k+1} 5^k. \tag{3.27}$$

It does not seem to be intuitively obvious that the sums should be divisible by 2^{n-1}, nor that the ratio of the sums should approach $\sqrt{5}$ as $n \to \infty$. It takes nothing away from the romance of Fibonacci numbers to observe that many of their properties are similar to those satisfied by *any* sequence satisfying the second order linear recurrence; $a(n) = \alpha a(n-1) + \beta a(n-2)$. This is especially true when $\beta = \pm 1$ (so the roots of the characteristic equation are reciprocals or nearly so), and when $a(0) = 0, a(1) = 1$.

3.4. **Two basic Stern recurrences.** In this section, we make a first pass at answering two basic questions about the Stern sequence: how many n have the property that $3 \mid s(n)$, and what is the behavior of $\sum s(n)^2$?

Recalling (1.76) and (1.77), $A(3,0)$ denotes the set of n for which $3 \mid s(n)$, $U(r;3,0)$ is the number of elements of $A(3,0)$ in I_r and $T(n;3,0)$ is the number of elements of $A(3,0)$ which are $\le n$. We showed in Theorem 1.7 that $0 < n \in A(3,0)$ if and only if $2n, 8n \pm 5, 8n \pm 7 \in A(3,0)$. We set $a_r = U(r;3,0)$ for short.

We can easily compute the first few values of $a_r := U(r;3,0)$; namely, $a_0 = a_1 = 0, a_2 = a_3 = a_4 = 2, a_5 = 10, a_6 = 18$.

Theorem 3.3. *If $r \ge 3$, then*

$$a_r = a_{r-1} + 4a_{r-3}. \tag{3.28}$$

Proof. Since $s(2^r) = 1$, no power of 2 is in $A(3,0)$. By (1.81), the number of n in $A(3,0)$ in the five "covering congruences" of 0 (mod 2), 5 (mod 8), -5 (mod 8), 7 (mod 8), -7 (mod 8) is equal to $a_{r-1}, a_{r-3}, a_{r-3}, a_{r-3}, a_{r-3}$ respectively. □

The characteristic equation of (3.28) is $z^3 - z^2 - 4 = (z-2)(z^2+z+2) = (z-2)(z-\mu)(z-\bar{\mu})$, where

$$\mu = \frac{-1+\sqrt{7}i}{2} \approx -.5 + 1.323i, \quad \bar{\mu} = \frac{-1-\sqrt{7}i}{2} \approx -.5 - 1.323i. \tag{3.29}$$

It follows that $a_r = c_1 2^r + c_2 \mu^r + c_3 \bar{\mu}^r$, where

$$\begin{aligned} 0 &= c_1 + c_2 + c_3 \\ 0 &= 2c_1 + \mu c_2 + \bar{\mu} c_3 \\ 2 &= 4c_1 + \mu^2 c_2 + \bar{\mu}^2 c_3. \end{aligned} \tag{3.30}$$

Since $|\mu| = |\bar{\mu}| = \sqrt{2}$, the asymptotic growth of a_r is determined by c_1. Although (3.30) is easy to solve, there is a trick to computing c_1 directly. Observe that μ and $\bar{\mu}$ are the roots of $z^2 + z + 2 = 0$. Thus, if we multiply the rows above by 2, 1, and 1, successively and add, we find $2 = 8c_1$, so $c_1 = \frac{1}{4}$. A routine computation, the rest of which we omit, shows that

$$a_r = \frac{1}{4} \cdot 2^r + \left(\frac{-7+5\sqrt{7}i}{56} \right) \mu^r + \left(\frac{-7-5\sqrt{7}i}{56} \right) \bar{\mu}^r. \tag{3.31}$$

It follows by a routine (but easy-to-get-wrong) computation using $\mu\bar{\mu} = 2$ that:

$$\begin{aligned} T(2^r;3,0) = \sum_{k=0}^{r-1} a_k &= \frac{2^r}{4} + \left(\frac{i}{4\sqrt{7}} \right) \cdot (\bar{\mu}\mu^r - \mu\bar{\mu}^r) + \frac{1}{2} \\ &= \frac{2^r}{4} + \left(\frac{i}{2\sqrt{7}} \right) \cdot (\mu^{r-1} - \bar{\mu}^{r-1}) + \frac{1}{2}. \end{aligned} \tag{3.32}$$

As $|\mu| = |\bar{\mu}| = \sqrt{2}$, it follows that $|T(2^r; 3, 0) - \frac{2^r}{4}| = \mathcal{O}(2^{r/2})$. In Theorem 3.12 we prove the stronger estimate that $T(n; 3, 0) = \frac{n}{4} + \mathcal{O}(n^{1/2})$.

Another way to look at these equations is to write

$$(\sqrt{2}) \cdot \frac{-1 \pm \sqrt{7}i}{2\sqrt{2}} = \sqrt{2}e^{\pm i\alpha}, \quad \frac{-7 \pm 5\sqrt{7}i}{56} = \frac{1}{\sqrt{14}} \cdot e^{\pm i\beta}. \tag{3.33}$$

Then (3.31) becomes

$$a_r = \frac{1}{4} \cdot 2^r + \frac{2^{r/2}}{\sqrt{14}} \cdot \left(e^{i(r\alpha+\beta)} + e^{-i(r\alpha+\beta)}\right) = \frac{1}{4} \cdot 2^r + 2^{r/2}\sqrt{2/7}\cos(r\alpha + \beta). \tag{3.34}$$

Niven's Theorem states that if $\frac{\theta}{\pi}$ and $\cos(\theta)$ are both rational, then $2\cos(\theta) \in \mathbb{N}$. It follows from (3.33) that $\cos(2\alpha) = \frac{3}{4}$, hence $\frac{\alpha}{2\pi}$ is irrational. It follows that the values of the sequence $(\cos(r\alpha + \beta))$ are dense in $[-1, 1]$; the coefficient of the error term in (3.34) thus gets arbitrarily close to $\sqrt{2/7} \approx .5345$.

We turn to the second question. We saw in (1.33) that $\sum_{n=2^r}^{2^{r+1}-1} s(n) = 3^r$. What can one say about $\sum_{n=2^r}^{2^{r+1}-1} s(n)^2$? It is helpful to make a more general definition. For integers $u, v \geq 0$, let

$$m_{u,v}(r) := \sum_{n=2^r}^{2^{r+1}-1} s(n)^u s(n+1)^v. \tag{3.35}$$

A quick lemma uses the row-reflection property (1.19):

Lemma 3.4. *For all u, v, r, $m_{u,v}(r) = m_{v,u}(r)$.*

Proof. A reparameterization via $m + n = 3 \cdot 2^r - 1 = 2^{r+1} + 2^r - 1$ shows that

$$\begin{aligned} m_{u,v}(r) &= \sum_{n=2^r}^{2^{r+1}-1} s(n)^u s(n+1)^v = \sum_{m=2^{r+1}-1}^{2^r} s(3 \cdot 2^r - m - 1)^u s(3 \cdot 2^r - m)^v \\ &= \sum_{m=2^r}^{2^{r+1}-1} s((m+1)^*)^u s(m^*)^v = \sum_{m=2^r}^{2^{r+1}-1} s((m+1))^u s(m)^v = m_{v,u}(r). \end{aligned}$$

□

Lemma 3.5. *The following family of recurrences hold for $r \geq 0$:*

$$m_{u,v}(r+1) = \sum_{k=0}^{v} \binom{v}{k} m_{u+k,v-k}(r) + \sum_{\ell=0}^{u} \binom{u}{\ell} m_{u-\ell,v+\ell}(r). \tag{3.36}$$

Proof. The general reindexing

$$\sum_{n=2a}^{2b-1} f(n) = \sum_{n=a}^{b-1} f(2n) + \sum_{n=a}^{b-1} f(2n+1), \tag{3.37}$$

applied to $a = 2^r, b = 2^{r+1}$, implies that

$$m_{u,v}(r+1) = \sum_{n=2^r}^{2^{r+1}-1} s(2n)^u s(2n+1)^v + \sum_{n=2^r}^{2^{r+1}-1} s(2n+1)^u s(2n+2)^v$$
$$= \sum_{n=2^r}^{2^{r+1}-1} s(n)^u (s(n)+s(n+1))^v + \sum_{n=2^r}^{2^{r+1}-1} (s(n)+s(n+1))^u s(n+1)^v$$

Expansion by the binomial theorem leads to (3.36). □

Theorem 3.6. *For each integer p, the sequences $(m_{i,p-i}(r))$, $0 \le i \le p$, satisfy the same linear recurrence of order $\lfloor \frac{p}{2} \rfloor + 1$.*

Proof. There are $p+1$ sequences $(m_{i,p-i}(r))$, and (3.36) shows that they satisfy a matrix linear recurrence. But Lemma 3.4 shows that we can rewrite (3.36) so as to limit our attention to the $\lfloor \frac{p}{2} \rfloor + 1$ sequences $(m_{i,p-i}(r))$ with $i \ge p/2$. □

In particular, Lemmas 3.4 and 3.5 imply that $m_{2,0}(r) = m_{0,2}(r)$ and

$$\begin{gathered} m_{2,0}(r+1) = m_{2,0}(r) + m_{2,0}(r) + 2m_{1,1}(r) + m_{0,2}(r), \\ m_{1,1}(r+1) = m_{2,0}(r) + m_{1,1}(r) + m_{1,1}(r) + m_{0,2}(r) \\ \Longrightarrow \begin{pmatrix} m_{2,0}(r+1) \\ m_{1,1}(r+1) \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} m_{2,0}(r) \\ m_{1,1}(r) \end{pmatrix}. \end{gathered} \tag{3.38}$$

The characteristic polynomial of $\left[\begin{smallmatrix} 3 & 2 \\ 2 & 2 \end{smallmatrix}\right]$ is $\lambda^2 - 5\lambda + 2$, which has roots

$$\nu = \frac{5+\sqrt{17}}{2} \approx 4.562, \qquad \bar{\nu} = \frac{5-\sqrt{17}}{2} \approx .438. \tag{3.39}$$

The initial conditions are: $m_{2,0}(0) = s(1)^2 = 1$, $m_{2,0}(1) = s(2)^2 + s(3)^2 = 5$. After some computations, $m_{2,0}(r)$ simplifies to:

Theorem 3.7.

$$m_{2,0}(r) = \frac{1}{\sqrt{17}} \cdot (\nu^{r+1} - \bar{\nu}^{r+1}). \tag{3.40}$$

Since $\bar{\nu} \in (0,1)$, it follows that $m_{2,0}(r) = \lfloor \nu^{r+1}/\sqrt{17} \rfloor$. The formula for $m_{1,1}(r)$ can be found from the relation, $m_{1,1}(r+1) = m_{2,0}(r+1) - m_{2,0}(r)$. For later reference, we observe that $s(2^r) = s(2^{r+1}) = 1$ implies that

$$m_{u,0}(r) = m_{0,u}(r) = \sum_{n \in I_r}^{*} s(n)^u. \tag{3.41}$$

This is not true for $m_{u,v}(r)$ when $uv > 0$ because $s(2^r - 1) \ne s(2^{r+1} + 1)$.

3.5. **A summation technique.** Recall the "trapezoidal sum" from (1.28):

$$\sum_{n=a}^{b}{}^{*} f(n) = \sum_{n=a}^{b} f(n) - \frac{f(a)+f(b)}{2} = \sum_{n=a}^{b-1} \frac{f(n)+f(n+1)}{2};$$

these expressions are additive; c.f. (1.29).

For $2^r m \le n \le 2^r(m+1)$, $s(n)$ is easily expressible in terms of $s(m)$ and $s(m+1)$. For this reason, we define the sequence $S(f;m) = (S(f;m,r))$ by

$$S(f;m,r) := \sum_{n=2^r m}^{2^r(m+1)}{}^{*} f(n). \tag{3.42}$$

Lemma 3.8. *Suppose that there is a finite set of sequences $\{(a_j(r)) : 1 \le j \le e\}$ with the property that, for given f and each $m \in \mathbb{N}$, $S(f;m,r) = a_{j_m}(r)$ for some $j_m, 1 \le j_m \le e$. Then there exists $d \le e$ and c_ℓ, $1 \le \ell \le d$, so that, for all m:*

$$S(f;m,d) + \sum_{\ell=1}^{d} c_j S(f;m,d-\ell) = 0. \tag{3.43}$$

Proof. The $e+1$ e-tuples $v(r) := (a_1(r), \dots a_e(r))$, $0 \le r \le e$, are linearly dependent, so $\sum_{r=0}^{e} \lambda_r v(r) = 0$ for some non-zero λ_r. Let d be the largest r so that $\lambda_r \neq 0$ and then let $c_j = \lambda_{d-j}/\lambda_d$ for $0 \le j \le d$. □

When there is no ambiguity, for a sequence $(a(r))$, we define

$$Y(a;r) := a(r) + \sum_{\ell=1}^{d} c_j a(r-\ell) = 0, \tag{3.44}$$

so that (3.43) is simply $Y(S(f;m);d) = 0$.

Lemma 3.9. *For all (f,m,r), we have*

$$S(f;m,r+1) = S(f;2m,r) + S(f;2m+1,r). \tag{3.45}$$

Proof. This is an application of (1.29); simply reinterpret

$$\sum_{n=2^{r+1}m}^{2^{r+1}(m+1)}{}^{*} f(n) = \sum_{n=2^r(2m)}^{2^r(2m+1)}{}^{*} f(n) + \sum_{n=2^r(2m+1)}^{2^r(2m+2)}{}^{*} f(n). \tag{3.46}$$

□

Lemma 3.10. *If $Y(S(f,m);d) = 0$ for all m, then for all $r \ge d$,*

$$Y(S(f;m);r) = 0. \tag{3.47}$$

Proof. We prove (3.47) by induction on r; the base case is (3.43). Assuming that (3.47) holds, we find that

$$Y(S(f;m);r+1) = Y(S(f;2m);r) + Y(S(f;2m+1);r) \tag{3.48}$$

by repeated application of Lemma 3.9, and this establishes the inductive step. □

Theorem 3.11. *Suppose $Y(S(f,m);d) = 0$ for all m and for $t \in \mathbb{N}$, let*

$$A_t(r) = \sum_{n=0}^{2^r t}{}^{*} f(n). \tag{3.49}$$

Then for $r \geq d$, the sequence $(A_t(r))$ satisfies

$$Y(A_t; r) = 0 \tag{3.50}$$

Proof. Suppose $t = 2^{r_1} + 2^{r_2} + \cdots + 2^{r_k}$, with $r_1 > r_2 > \cdots > r_k$, and let $N_0 = 0$ and $N_j = 2^{r_1} + \cdots + 2^{r_j} = N_{j-1} + 2^{r_j}$ for $j = 1, \ldots, k$, so that $t = N_k$. Further, for $1 \leq j \leq k$, let $M_j = 2^{-r_j} N_{j-1}$, so that $N_{j-1} = 2^{r_j} M_j$ and $N_j = 2^{r_j}(M_j + 1)$. Then

$$\begin{aligned} A_t(r) &= \sum_{n=0}^{2^r t}{}^{*} f(n) = \sum_{j=1}^{k} \sum_{n=2^r N_{j-1}}^{2^r N_j}{}^{*} f(n) \\ &= \sum_{j=1}^{k} \left(\sum_{n=2^{r+r_j} M_j}^{2^{r+r_j}(M_j+1)}{}^{*} f(n) \right) = \sum_{j=1}^{k} S(f; M_j, r + r_j). \end{aligned} \tag{3.51}$$

is a sum of sequences, each of which satisfies (3.50) by Lemma 3.10. □

The hypotheses of Theorem 3.11 might seem to be formidably strong, and they usually are, unless the function f being summed relates to the Stern sequence. For example, suppose $f(n) = s(n)$ itself, and write $f(m) = a$ and $s(m+1) = b$. Then

$$\begin{aligned} S(f; m, 0) &:= \frac{s(m)}{2} + \frac{s(m+1)}{2} = \frac{a+b}{2}, \\ S(f; m, 1) &:= \frac{s(2m)}{2} + s(2m+1) + \frac{s(2m+2)}{2} \\ &= \frac{a + 2(a+b) + b}{2} = \frac{3a+3b}{2}. \end{aligned} \tag{3.52}$$

That is, $S(f; m, 1) = 3S(f; m, 0)$, and so, as we've already seen in Lemma 1.3,

$$\sum_{n=0}^{2N}{}^{*} s(n) = 3 \sum_{n=0}^{N}{}^{*} s(n). \tag{3.53}$$

We will now apply Theorem 3.11 to the two situations of §3.4.

3.6. **The Stern sequence mod 3.** Suppose $f = \chi_{A(d,i)}$, so for $T \subseteq \mathbb{N}$, the expression $\sum_{n \in T} f(n)$ counts the number of n in T for which $s(n) \equiv i \pmod{d}$. By (1.12),

$$s(2^r m + k) = s(2^r - k)s(m) + s(k)s(m+1), \quad 0 \leq k \leq 2^r.$$

If $(s(m_1), s(m_1+1)) \equiv (s(m_2), s(m_2+1)) \pmod{d}$, then it follows that $s(2^r m_1 + k) \equiv s(2^r m_2 + k) \pmod{d}$ for $0 \leq k \leq 2^r$, so that

$$S(\chi_{A(d,i)}; m_1, r) = S(\chi_{A(d,i)}; m_2, r). \tag{3.54}$$

Since $s(m) \equiv s(m+1) \equiv 0 \pmod{d}$ is impossible, this means there are at most d^2-1 different sequences $(S(\chi_{A(d,i)}; m, r))$, parameterized by $(i \pmod{d}, j \pmod{d})$. The hypothesis of Lemma 3.8 is satisfied and so Theorem 3.11 applies.

We now apply this in detail to the case $(d, i) = (3, 0)$, $f = \chi = \chi_{A_{3,0}}$ and keep this notation for the rest of this section; we return to the general case in a later chapter.

There are, conceivably eight different cases, depending on the congruence classes mod 3; however, two observations reduce the number of different sequences: first, if $(s(m'), s(m'+1)) \equiv (s(m+1), s(m)) \pmod 3$, then $S(\chi; m, r) = S(\chi; m', r)$ by mirror symmetry; second, if $(s(m'), s(m'+1)) \equiv -(s(m), s(m+1)) \pmod 3$, then $s(2^r m' + k) \equiv -s(2^r m + k) \pmod 3$, so $\chi(2^r m' + k) = \chi(2^r m + k)$ and again, $S(\chi; m, r) = S(\chi; m', r)$.

Let $S_{ij}(r) = S(\chi; m, r)$ in the case that $(s(m), s(m+1)) \equiv (i, j) \pmod 3$. Thus, $S_{01}(r) = S_{02}(r) = S_{10}(r) = S_{20}(r)$, $S_{11}(r) = S_{22}(r)$ and $S_{12}(r) = S_{21}(r)$, so there are only three different sequences. A look at the initial conditions

(3.55)

$$
\begin{array}{lccccccccc}
S_{01}(r): & 0 & & & & & & & & 1 \\
 & 0 & & & & 1 & & & & 1 \\
 & 0 & & 1 & & 1 & & 2 & & 1 \\
 & 0 & 1 & 1 & 2 & 1 & 3 & 2 & 3 & 1 \\
S_{11}(r): & 1 & & & & & & & & 1 \\
 & 1 & & & & 2 & & & & 1 \\
 & 1 & & 3 & & 2 & & 3 & & 1 \\
 & 1 & 4 & 3 & 5 & 2 & 5 & 3 & 4 & 1 \\
S_{12}(r): & 1 & & & & & & & & 2 \\
 & 1 & & & & 3 & & & & 2 \\
 & 1 & & 4 & & 3 & & 5 & & 2 \\
 & 1 & 5 & 4 & 7 & 3 & 8 & 5 & 7 & 2
\end{array}
$$

shows that

(3.56)

$$
\begin{aligned}
(S_{01}(0), S_{11}(0), S_{12}(0)) &= (\tfrac{1}{2}, 0, 0), \\
(S_{01}(1), S_{11}(1), S_{12}(1)) &= (\tfrac{1}{2}, 0, 1), \\
(S_{01}(2), S_{11}(2), S_{12}(2)) &= (\tfrac{1}{2}, 2, 1), \\
(S_{01}(3), S_{11}(3), S_{12}(3)) &= (\tfrac{5}{2}, 2, 1).
\end{aligned}
$$

It follows, unsurprisingly in view of Theorem 3.3, that

$$S_{ij}(3) = S_{ij}(1) + 4S_{ij}(0). \tag{3.57}$$

This derivation of the recurrence uses no combinatorial information; instead, its existence is forced by dimensional considerations. An alternative approach to this method would apply Lemma 3.9 to observe that $S_{ij}(r+1) = S_{ik}(r) + S_{kj}(r)$, where $i + j \equiv k \pmod 3$. (This matrix recurrence is the technique we will use for general bases in a later chapter.)

In the notation of §3.4, we have $S_{11}(r) = a_r$, $S_{12}(r) = \frac{1}{2}a_{r+1}$ and $S_{01}(r) = T(2^r;3,0) = \frac{1}{4}a_{r+2}$, and further

$$\begin{aligned} \left|S_{11}(r) - \tfrac{1}{4}\cdot 2^r\right| &\le \sqrt{\tfrac{2}{7}}\cdot(\sqrt{2})^r, \\ \left|S_{12}(r) - \tfrac{1}{4}\cdot 2^r\right| &\le \sqrt{\tfrac{1}{7}}\cdot(\sqrt{2})^r, \\ \left|S_{01}(r) - \tfrac{1}{4}\cdot 2^r\right| &\le \sqrt{\tfrac{1}{14}}\cdot(\sqrt{2})^r. \end{aligned} \tag{3.58}$$

Thus, it follows that for all (m,r),

$$\left|S(\chi;m,r) - \frac{1}{4}\cdot 2^r\right| \le \sqrt{\tfrac{2}{7}}\cdot(\sqrt{2})^r. \tag{3.59}$$

Theorem 3.12.

$$\left|T(N;3,0) - \frac{N}{4}\right| = \mathcal{O}(N^{1/2}). \tag{3.60}$$

Proof. First observe that

$$T(N;3,0) = \sum_{n=0}^{2^N}{}^{*}\,\chi(n) + \frac{\chi(0)+\chi(N)}{2},$$

so the difference between the two is either $\frac{1}{2}$ or 1, depending on whether $N \in A(3,0)$. This will not affect the asymptotics. Next, suppose that $N = 2^{r_1} + \cdots + 2^{r_k}$. Then

$$\sum_{n=0}^{N}{}^{*}\,\chi(n) = \sum_{j=1}^{k} S(\chi;M_j,r_j). \tag{3.61}$$

Therefore,

$$\begin{aligned} \left|T(3,0;N) - \frac{N}{4}\right| &\le 1 + \left|\sum_{n=0}^{N}{}^{*}\,\chi(n) - \frac{N}{4}\right| \\ &\le 1 + \left|\sum_{j=1}^{k} S(\chi;M_j,r_j) - \frac{1}{4}\sum_{j=1}^{k} 2^{r_j}\right| \le 1 + \sum_{j=1}^{k}\left|S(\chi;M_j,r_j) - \frac{1}{4}\cdot 2^{r_j}\right| \\ &\le 1 + \sum_{j=1}^{k}\sqrt{\tfrac{2}{7}}\cdot(\sqrt{2})^{r_j} \le 1 + \sqrt{\tfrac{2}{7}}\left(\sum_{\ell=0}^{r_1}(\sqrt{2})^{\ell}\right) = 1 + \sqrt{\tfrac{2}{7}}\cdot\frac{(\sqrt{2})^{r_1+1}-1}{\sqrt{2}-1} \\ &< 1 + \sqrt{\tfrac{2}{7}}(\sqrt{2}+1)(\sqrt{2})(N^{1/2}-1) = \frac{2\sqrt{2}+2}{\sqrt{7}}\cdot N^{1/2} < 2N^{1/2}. \end{aligned} \tag{3.62}$$

□

3.7. **How fast does the Stern sequence grow?** We return briefly to Theorem 3.7. Using the general methods, we can obtain the recurrence more easily. Writing $s(m) = a$ and $s(m+1) = b$, with $f(n) = s(n)^2$, we have

$$\begin{aligned} S(f;m,0) &= \frac{f(m)}{2} + \frac{f(m+1)}{2} = \frac{a^2+b^2}{2}, \\ S(f;m,1) &= \frac{f(2m)}{2} + f(2m+1) + \frac{f(2m+2)}{2} \\ &= \frac{a^2 + 2(a+b)^2 + b^2}{2} = \frac{3a^2+4ab+3b^2}{2}, \\ S(f;m,2) &= \frac{f(4m)}{2} + f(4m+1) + f(4m+2) + f(4m+3) + \frac{f(4m+4)}{2} \\ &= \frac{3a^2+4ab+3b^2}{2} + (2a+b)^2 + (a+2b)^2 = \frac{13a^2+20ab+13b^2}{2}. \end{aligned} \tag{3.63}$$

It may be readily verified that

$$S(f;m,2) - 5S(f;m,1) + 2S(f,m,0) = 0. \tag{3.64}$$

Since $S(f;1,r) = m_{2,0}(r)$, this is a much faster derivation of Theorem 3.7; however, you don't get the recurrence for $g(N)$, which is *not* a sum of this kind.

We now discuss, without tremendous detail,

$$m_{k,0}(r) := \sum_{n=2^r}^{2^{r+1}\,*} s(n)^k = S(s(n)^k;1,r). \tag{3.65}$$

As above, we suppose that $s(m) = a$ and $s(m+1) = b$, so that

$$\begin{aligned} S(f;m,0) &= \frac{1}{2}\big(a^k + b^k\big); \\ S(f;m,1) &= S(f;m,0) + (a+b)^k; \\ S(f;m,2) &= S(f;m,1) + (2a+b)^k + (a+2b)^k. \end{aligned} \tag{3.66}$$

As we have already seen, when $k = 1$, $S(f;m,1) = 3S(f;m,0)$ and $m_{1,0}(r) = 3^r$. When $k = 3$, there is an unusually pleasant recurrence:

$$S(f;m,2) - 7S(f;m,1) = (2a+b)^3 + (a+2b)^3 - 6(a+b)^3 - 3(a^3+b^3) = 0.$$

It follows that $m_{3,0}(r+2) = 7m_{3,0}(r+1)$ for $r \geq 0$, and that

$$m_{3,0}(0) = 1, \qquad m_{3,0}(r) = 9 \cdot 7^{r-1}, \qquad r \geq 1. \tag{3.67}$$

Note that this second-order recurrence has one root equal to zero and another way to express (3.67) would be as $m_{3,0}(r) = \frac{9}{7} \cdot 7^r - \frac{2}{7} \cdot 0^r$. Alternatively, Lemma 3.5 implies

that $m_{3,0}(r) = m_{0,3}(r)$, $m_{2,1}(r) = m_{1,2}(r)$ and

$$\begin{aligned} m_{3,0}(r+1) &= 2m_{3,0}(r) + 3m_{2,1}(r) + 3m_{1,2}(r) + m_{0,3}(r) \\ m_{2,1}(r+1) &= m_{3,0}(r) + 2m_{2,1}(r) + 2m_{1,2}(r) + m_{0,3}(r) \\ \implies \begin{pmatrix} m_{3,0}(r+1) \\ m_{2,1}(r+1) \end{pmatrix} &= \begin{pmatrix} 3 & 6 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} m_{3,0}(r) \\ m_{2,1}(r) \end{pmatrix}. \end{aligned} \tag{3.68}$$

The characteristic polynomial of $\left(\begin{smallmatrix} 3 & 6 \\ 2 & 4 \end{smallmatrix}\right)$ is $\lambda^2 - 7\lambda$. The cases for $k = 4, 5$ are deferred to the solutions to the second set of exercises.

Is there a point to this? We might ask the vague question: how large is $s(n)$? Certainly individual values vary quite a bit: if $2^r \le n \le 2^{r+1}$, $1 \le s(n) \le F_{r+2}$, as we have seen. However, the results on sums of powers suggest that there might be some regular behavior. Define the *t-th power mean* for $2^r \le n \le 2^{r+1}$:

$$M(r;t) := \left(\frac{1}{2^r} \sum_{n=2^r}^{2^{r+1}} {}^{*} s(n)^t \right)^{1/t}. \tag{3.69}$$

We have already seen that $M(r,1) = \left(\frac{3}{2}\right)^r$, which suggests the definition

$$L(r;t) := M(r,1)^{1/r}. \tag{3.70}$$

In this notation, $L(r;1) = \frac{3}{2}$. This subject ties in with some traditional analysis, and a classical slick inequality proof.

Theorem 3.13. *If $x_k > 0$ for $1 \le k \le n$, then the function*

$$M(t) := \left(\frac{1}{n} \sum_{k=1}^{n} x_k^t \right)^{1/t} \tag{3.71}$$

is increasing for $t \ge 0$, and $\lim_{t\to\infty} M(t) = \max x_k$.

Proof. We first consider the auxiliary function

$$\Phi(t) := \log \left(\sum_{k=1}^{n} x_k^t \right), \tag{3.72}$$

and commit calculus, finding by routine computation that

$$\Phi''(t) = \frac{\left(\sum_{k=1}^{n} x_k^t \right) \left(\sum_{k=1}^{n} x_k^t (\log x_k)^2 \right) - \left(\sum_{k=1}^{n} x_k^t \log x_k \right)^2}{\left(\sum_{k=1}^{n} x_k^t \right)^2}. \tag{3.73}$$

The numerator of (3.73) is non-negative by Cauchy-Schwartz, so Φ is convex for all t. This implies that

$$\Psi(t) := \frac{\Phi(t) - \Phi(0)}{t}$$

is an increasing function for $t \geq 0$, as is $e^{\Psi(t)}$. But

$$\Psi(t) = \frac{\log\left(\sum_{k=1}^n x_k^t\right) - \log\left(\sum_{k=1}^n x_k^0\right)}{t} = \frac{1}{t}\log\left(\frac{1}{n}\sum_{k=1}^n x_k^t\right) \tag{3.74}$$
$$\implies e^{\Psi(t)} = M(t).$$

If $M = \max x_k$, then $M^t \leq \sum_{k=1}^n x_k^t \leq nM^t$, so $n^{-1/t}M \leq M(t) \leq M$. □

For example,

$$\lim_{t\to\infty} L(r,t) = F_{r+2}^{1/r}.$$

Lemma 3.14. *Suppose*

$$a_r = c\lambda^r + \sum_{j=1}^t g_j(r)\lambda_j^r, \tag{3.75}$$

where $c > 0$, $g_j(r)$ is a polynomial in r and $\lambda > |\lambda_j|$ is a positive real. Then

$$\lim_{r\to\infty} a_r^{1/r} = \lambda. \tag{3.76}$$

Proof. Since $a_r = \lambda^r b_r$, where $\lim b_r = c > 0$, and since $c^{1/r} \to 1$, the proof is immediate. □

It follows from Lemma 3.14 that

$$\lim_{r\to\infty}\left(\lim_{t\to\infty} L(r,t)\right) = \lim_{r\to\infty} F_{r+2}^{1/r} = \Phi = \frac{1+\sqrt{5}}{2} \approx 1.6180. \tag{3.77}$$

A more interesting, and open, question, is the computation of

$$\lim_{t\to\infty}\left(\lim_{r\to\infty} L(r,t)\right). \tag{3.78}$$

which, based on the lemma, is closely related to the behavior of the recurrences that we've found. In fact, taking into account earlier equations and the numerical values of the roots not presented here explicitly (for $t = 4, 5, 6, 7$, and providing an uninteresting hint to the exercises), we have

$$\begin{gathered}\lim_{r\to\infty} L(r,1) = \tfrac{3}{2} = 1.5, \quad \lim_{r\to\infty} L(r,2) = \left(\tfrac{5+\sqrt{17}}{4}\right)^{1/2} \approx 1.5102,\\ \lim_{r\to\infty} L(r,3) = \left(\tfrac{7}{2}\right)^{1/3} \approx 1.5183, \quad \lim_{r\to\infty} L(r,4) \approx 1.5249,\\ \lim_{r\to\infty} L(r,5) \approx 1.5305, \lim_{r\to\infty} L(r,6) \approx 1.5354, \lim_{r\to\infty} L(r,7) \approx 1.5396.\end{gathered} \tag{3.79}$$

The limiting behavior is far from clear.

3.8. **Sylvester's Theorem.** The material in this section is mostly adapted from the papers *On the length of binary forms*, to appear in Quadratic and Higher Degree Forms, (K. Alladi, M. Bhargava, D. Savitt, P. Tiep, eds.), Developments in Math. Springer, New York, http://arxiv.org/pdf/1007.5485.pdf. and *Sums of even powers of real linear forms*, Mem. Amer. Math. Soc., Volume 96, Number 463, March, 1992 (MR 93h.11043), both of which can be downloaded from my website.

We give an application of linear recurrences with no direct ties to Stern sequences, but which resonates with various number theoretic questions: namely, the representation of binary forms as a sum of powers of linear forms. The main result was proved by J. J. Sylvester in 1851.

The representation of quadratic forms as a sum of squares of linear forms is well understood and a standard part of linear algebra. This is less so for higher powers of linear forms, even though the simplest case (two variables) has been understood for more than 150 years.

Consider also the classical question of quadrature. Suppose $S \subseteq \mathbb{R}^n$ and measure μ are given. A *quadrature formula of strength m for $(S, d\mu)$* is an exact formula

$$\int_S f \, d\mu = \sum_{k=1}^{r} \lambda_k f(t_k), \quad f \in \mathbb{R}[x_1, \ldots, x_n], \quad \deg f \le m. \tag{3.80}$$

In this section we take the case of $S = [-1, 1] \subseteq \mathbb{R}^1$ and Lebesgue measure: (3.80) is a quadrature formula of strength m provided it holds for $f(t) = t^i$, $0 \le i \le m$, so

$$\frac{1 - (-1)^{m+1}}{m+1} = \int_{-1}^{1} t^i \, dt = \sum_{k=1}^{r} \lambda_k t_k^i, \qquad 0 \le i \le m. \tag{3.81}$$

If $r < m$, then (3.81) says that the successive "moments" must satisfy an r-th order linear recurrence. We can turn (3.81) into a single equation by constructing the appropriate generating function:

$$\begin{aligned} \int_{-1}^{1} (x + ty)^m \, dt &= \sum_{i=0}^{m} \binom{m}{i} \left(\int_{-1}^{1} t^i \, dt \right) x^{m-i} y^i \\ &= \sum_{i=0}^{m} \binom{m}{i} \left(\sum_{k=1}^{r} \lambda_k t_k^i \right) x^{m-i} y^i = \sum_{k=1}^{r} \lambda_k (x + t_k y)^m. \end{aligned} \tag{3.82}$$

In other words, a quadrature formula on an interval in $\mathbb{R}$ is the same as a representation of a binary forms of degree m as a sum of m-th powers of linear forms.

We consider binary *m-ic* forms with complex coefficients; that is, homogeneous polynomials of degree m in two variables, written as:

$$f(x, y) = \sum_{i=0}^{m} \binom{m}{i} a_i x^{m-i} y^i. \tag{3.83}$$

(It is both customary and convenient to factor the binomial coefficient out in the expression.) The *length* or *rank* of f is the smallest integer r with the property that

there exists an expression:

$$f(x,y) = \sum_{k=1}^{r} \gamma_k(\alpha_k x + \beta_k y)^m. \tag{3.84}$$

One quick remark: if r in (3.84) is minimal, then the linear forms $\{\alpha_k x + \beta_k y\}$ will be pairwise non-proportional; otherwise, two could be combined. Under this restriction, we say that (3.84) is an *honest* representation.

If (3.84) is honest, then $\alpha_k = 0$ for at most one zero; hence after writing $c_k = \gamma_k \alpha_k^m$ and $\lambda_k = \beta_k/\alpha_k$,

$$f(x,y) = \sum_{k=1}^{r} c_k(x + \lambda_k y)^m, \quad \text{or} \quad f(x,y) = \sum_{k=1}^{r-1} c_k(x + \lambda_k y)^m + c_r y^m. \tag{3.85}$$

A comparison with (3.83) shows that (3.85) is equivalent to:

$$\begin{gathered} a_i = \sum_{k=1}^{r} c_k \lambda_k^i, \quad 0 \le i \le m \qquad \text{or} \\ a_i = \sum_{k=1}^{r-1} c_k \lambda_k^i \quad (0 \le i \le m-1), \quad a_m = \sum_{k=1}^{r-1} c_k \lambda_k^m + c_r \end{gathered} \tag{3.86}$$

Theorem 3.15 (Sylvester)**.** *Suppose f is given by* (3.83) *and suppose*

$$h(x,y) = \sum_{t=0}^{r} c_t x^{r-t} y^t = \prod_{j=1}^{r} (-\beta_j x + \alpha_j y) \tag{3.87}$$

is a given product of pairwise distinct linear factors. Then there exist $\lambda_k \in \mathbb{C}$ so that (3.84) *holds if and only if*

$$\begin{pmatrix} a_0 & a_1 & \cdots & a_r \\ a_1 & a_2 & \cdots & a_{r+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m-r} & a_{m-r+1} & \cdots & a_m \end{pmatrix} \cdot \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_r \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}; \tag{3.88}$$

that is, if and only if

$$\sum_{t=0}^{r} a_{\ell+t} c_t = 0, \qquad \ell = 0, 1, \ldots, m-r. \tag{3.89}$$

If (3.88) holds, we say that h, as defined by (3.87) is a *Sylvester form* for p. Sylvester's Theorem in practice involes writing the successive "Hankel" matrices in (3.88) with increasing r until there exists one with a non-zero null vector $(c_0, \ldots, c_r)^t$ so that the resulting Sylvester form has distinct factors. Afterwards, the computation of the λ_k is routine.

Proof. First suppose that (3.84) holds. Then for $0 \le i \le m$,

$$a_i = \sum_{k=1}^r \lambda_k \alpha_k^{m-i} \beta_k^i \implies \sum_{t=0}^r a_{\ell+t} c_t = \sum_{k=1}^r \sum_{t=0}^r \lambda_k \alpha_k^{m-\ell-t} \beta_k^{\ell+t} c_t$$

$$= \sum_{k=1}^r \lambda_k \alpha_k^{m-\ell-r} \beta_k^\ell \sum_{t=0}^r \alpha_k^{r-t} \beta_k^t c_t = \sum_{k=1}^r \lambda_k \alpha_k^{m-\ell-r} \beta_k^\ell \; h(\alpha_k, \beta_k) = 0.$$

Now suppose that (3.85) holds and suppose first that $c_r \neq 0$. We may assume without loss of generality that $c_r = 1$ and that $\alpha_j = 1$ in (3.87), so that the β_j's are distinct. Define the *infinite* sequence $(\tilde{a}_i)$, $i \ge 0$, by:

$$\tilde{a}_i = a_i \quad \text{if} \quad 0 \le i \le r-1; \qquad \tilde{a}_{r+\ell} = -\sum_{t=0}^{r-1} \tilde{a}_{t+\ell} c_t \quad \text{for} \quad \ell \ge 0; \tag{3.90}$$

so that $(\tilde{a}_i)$ satisfies (3.86) and extends the finite sequence $(a_0, \ldots, a_m)$:

$$\tilde{a}_i = a_i \quad \text{for} \quad i \le m. \tag{3.91}$$

Theorem 3.2 now implies that there exist λ_k so that for all i,

$$\tilde{a}_i = \sum_{k=1}^r \lambda_k \beta_k^i. \tag{3.92}$$

In particular,

$$f(x,y) = \sum_{i=0}^m \binom{m}{i} a_j x^{m-i} y^i = \tag{3.93}$$
$$\sum_{k=1}^r \lambda_k \sum_{i=0}^m \binom{m}{i} \beta_k^i x^{m-i} y^i = \sum_{k=1}^r \lambda_k (x + \beta_k y)^m,$$

as claimed in (3.84).

If $c_r = 0$, then $c_{r-1} \neq 0$, because h has distinct factors. We may proceed as before, replacing r by $r-1$ and taking $c_{r-1} = 1$, so that (3.87) becomes

$$h(x,y) = \sum_{t=0}^{r-1} c_t x^{r-t} y^t = x \prod_{j=1}^{r-1} (y - \beta_j x). \tag{3.94}$$

Since $c_r = 0$, the system (3.88) can be rewritten as

$$\begin{pmatrix} a_0 & a_1 & \cdots & a_{r-1} \\ a_1 & a_2 & \cdots & a_r \\ \vdots & \vdots & \ddots & \vdots \\ a_{d-r} & a_{m-r+1} & \cdots & a_{m-1} \end{pmatrix} \cdot \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{r-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

We may now argue as before, except that (3.91) becomes

$$\tilde{a}_i = a_i \quad \text{for} \quad i \le m-1, \qquad a_m = \tilde{a}_m + \lambda_m \tag{3.95}$$

for some λ_m, and (3.93) becomes

$$f(x,y) = \sum_{i=0}^{m} \binom{m}{i} a_j x^{m-i} y^i = \tag{3.96}$$
$$\lambda_r y^m + \sum_{k=1}^{r-1} \lambda_k \sum_{i=0}^{m} \binom{m}{i} \beta_k^i x^{m-i} y^i = \lambda_r y^m + \sum_{k=1}^{r-1} \lambda_k (x+\beta_k y)^m,$$

By (3.94), (3.96) meets the description of (3.84), completing the proof. □

In 1886, Gundelfinger studied the case where the Sylvester form h has repeated factors. The factor $(-\beta x+\alpha y)^\ell$ of h corresponds to a summand $q(x,y)(\alpha x+\beta y)^{d+1-\ell}$ in f, where q is an arbitrary form of degree $\ell-1$.

The classical application of Sylvester's Theorem is to "canonical forms". If $m = 2s-1$ and $r=s$, then the matrix in (3.88) has dimensions $s \times (s+1)$ and so has a non-trivial null-vector; for a "general" f, the resulting form h has distinct factors, and so a general binary form of degree $2s-1$ has a representation as a sum of s $2s-1$-st powers of linear forms, which is unique up to a permutation of the summands. If $d = 2s$ and $r = s$, then the matrix in (3.88) is square, and since its determinant is not 0 in general, there is no corresponding h. However, for general f, for any α, there there exists $\lambda_0 = \lambda_0(\alpha)$ so that the Hankel matrix for $f - \lambda_0(x+\alpha y)^{2s}$ *does* have a non-trivial null vector and is a sum of s $2s$-th powers, hence general f is a sum of $s+1$ $2s$-th powers in infinitely many ways.

Without going into details, one can also define the *K-length* of a form f for any subfield $K \subseteq \mathbb{C}$ which contains the coefficients of f, and Sylvester's Theorem can be adapted to that case as well. Here is one example:

$$H(x,y) = 3x^5 - 20x^3y^2 + 10xy^4 = \binom{5}{0} \cdot 3\ x^5 + \binom{5}{1} \cdot 0\ x^4 y$$
$$+\binom{5}{2} \cdot (-2)\ x^3y^2 + \binom{5}{3} \cdot 0\ x^2y^3 + \binom{5}{4} \cdot 2\ xy^4 + \binom{5}{5} \cdot 0\ y^5;$$
$$\begin{pmatrix} 3 & 0 & -2 & 0 \\ 0 & -2 & 0 & 2 \\ -2 & 0 & 2 & 0 \end{pmatrix} \cdot \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \iff (c_0,c_1,c_2,c_3) = r(0,1,0,1).$$

Thus, H has a *unique* Sylvester form of degree 3: $h(x,y) = y(x^2+y^2)$, which factors as $y(y-ix)(y+ix)$. Accordingly, there exist $\lambda_k \in \mathbb{C}$ so that

$$H(x,y) = \lambda_1 x^5 + \lambda_2 (x+iy)^5 + \lambda_3 (x-iy)^5.$$

It may be checked that $\lambda_1 = \lambda_2 = \lambda_3 = 1$; this is how H was constructed.

To find representations for H of length 4, we consider (3.85) for H with $r=4$:

$$H_4(H) \cdot (c_0,c_1,c_2,c_3,c_4)^t = (0,0)^t \iff 3c_0 - 2c_2 + 2c_4 = -2c_1 + 2c_3 = 0$$
$$\iff (c_0,c_1,c_2,c_3,c_4) = r_1(2,0,3,0,0) + r_2(0,1,0,1,0) + r_3(0,0,1,0,1),$$

hence $h(x,y) = r_1x^2(2x^2+3y^2) + y(x^2+y^2)(r_2x+r_3y)$. Given a field K, it is unclear whether there exist $\{r_\ell\}$ so that h splits into distinct factors over K. We have found such $\{r_\ell\}$ for small imaginary quadratic fields. For example, the choice $(r_1,r_2,r_3) = (1,0,2)$ gives $h(x,y) = (2x^2+y^2)(x^2+2y^2)$ and

$$24H(x,y) = 4(x+\sqrt{-2}y)^5 + 4(x-\sqrt{-2}y)^5 + (2x+\sqrt{-2}y)^5 + (2x-\sqrt{-2}y)^5.$$

Similarly, $(r_1,r_2,r_3) = (2,0,9)$ and $(2,0,-5)$ give $h(x,y) = (x^2+3y^2)(4x^2+3y^2)$ and $(x^2-y^2)(4x^2+5y^2)$, leading to representations for H of length 4 over $\mathbb{Q}(\sqrt{-3})$ and $\mathbb{Q}(\sqrt{-5})$. The simplest such representation we have found for $\mathbb{Q}(\sqrt{-6})$ uses $(r_1,r_2,r_3) = (8450,0,-104544)$ and

$$h(x,y) = (5x+12y)(5x-12y)(6\cdot 13^2x^2 + 33^2y^2).$$

We believe, but have not proved, that examples such as these exist for every imaginary quadratic field. A different theorem of Sylvester, from 1864, implies that H has *no* representation as a sum of fewer than five *real* fifth powers of linear forms.

As another example with number theory applications, consider the representations of $(xy)^k$ as a sum of $2k$-th powers of linear forms. The square Hankel matrix for $f(x,y) = \binom{2k}{k}x^ky^k$ has 1's on the NE-SW diagonal, and so is non-singular. Thus there is no representation of f as a sum of k $2k$-th powers of linear forms.

Let $\zeta_m = e^{2\pi i/m}$. It is easy to see that $\sum_{j=0}^{m-1}\zeta_m^{rj} = 0$ unless $m \mid r$, in which case it equals m. It turns out that the full set of minimal representations of $f(x,y) = \binom{2k}{k}x^ky^k$ as a sum of $(k+1)$ $2k$-th powers is

$$(k+1)\binom{2k}{k}x^ky^k = \sum_{j=0}^{k}(\zeta_{2k+2}^j wx + \zeta_{2k+2}^{-j}w^{-1}y)^{2k}, \qquad 0 \neq w \in \mathbb{C}. \tag{3.97}$$

Evaluate the right-hand side of (3.97) by expanding the powers:

$$\begin{aligned}\sum_{j=0}^{k}(\zeta_{2k+2}^j wx + \zeta_{2k+2}^{-j}w^{-1}y)^{2k} &= \sum_{j=0}^{k}\sum_{t=0}^{2k}\binom{2k}{t}\zeta_{2k+2}^{j(2k-t)-jt}w^{(2k-t)-t}x^{2k-t}y^t \\ &= \sum_{t=0}^{2k}\binom{2k}{t}w^{2k-2t}x^{2k-t}y^t\left(\sum_{j=0}^{k}\zeta_{k+1}^{j(k-t)}\right).\end{aligned} \tag{3.98}$$

Since the only multiple of $k+1$ in the set $\{k-t : 0 \le t \le 2k\}$ occurs for $t=k$, (3.98) reduces to the left-hand side of (3.97). The representations in (3.97) arise because the null-vectors of the resulting $(k-1)\times(k+1)$ Hankel matrix can only be $(c_0,0,\dots,0,c_{k+1})^t$ and $c_0x^{k+1} + c_{k+1}y^{k+1}$ is a Sylvester form when $c_0c_{k+1} \neq 0$. We state without proof that by making the change of variables in $(x,y) \mapsto (x-iy, x+iy)$ in (3.98) we obtain the expressions:

$$\binom{2k}{k}(x^2+y^2)^k = \frac{1}{k+1}\sum_{j=0}^{k}\left(\cos(\tfrac{j\pi}{k+1}+\theta)x + \sin(\tfrac{j\pi}{k+1}+\theta)y\right)^{2k}, \quad \theta \in \mathbb{C}. \tag{3.99}$$

When θ is real, these are related to the "Hilbert identities" used in solving the Waring problem. Even for real θ, the earliest instance of (3.99) in the literature seems to be by Friedman from 1957. Representations of $(x_1^2 + \cdots + x_n^2)^k$ as a real sum of $2k$-th powers and can be identified with quadrature formulas of strength $2k+1$ on the $S^{n-1} \subseteq \mathbb{R}^n$. In this sense, (3.99) can be traced back to work of Mehler from 1864.

We finish by revisiting quadrature formulas on $[-1, 1]$. For strength 3, (3.82) becomes:

$$2x^3 + 3 \cdot \tfrac{2}{3} xy^2 = \sum_{k=1}^{r} \lambda_k (x + t_k y)^3.$$

Sylvester's Theorem with $r = 2$ yields

$$\begin{pmatrix} 2 & 0 & 2/3 \\ 0 & 2/3 & 0 \end{pmatrix} \cdot \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff (c_0, c_1, c_2) = r(1, 0, -3).$$

Thus, $h(x, y) = x^2 - 3y^2$ and there exist λ_k so that $2x^3 + 3 \cdot \frac{2}{3} xy^2 = \lambda_1(\sqrt{3}x + y)^3 + \lambda_2(\sqrt{3}x - y)^3$. Cleaning up, we find the Gaussian quadrature formula of strength 3:

$$2x^3 + 2xy^2 = (x + \gamma y)^3 + (x - \gamma y)^3 \iff \int_{-1}^{1} f(t)\, dt = f(\gamma) + f(-\gamma); \quad \gamma = \sqrt{\tfrac{1}{3}}.$$

For strength 4, the matrix

$$\begin{pmatrix} 2 & 0 & 2/3 \\ 0 & 2/3 & 0 \\ 2/3 & 0 & 2/5 \end{pmatrix}$$

is non-singular, so there are no 2-point quadrature formulas of strength 3. But

$$\begin{pmatrix} 2 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 2/5 \end{pmatrix} \cdot \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

leads to $h(x, y) = rx(3x^2 - 2y^2) + sy(5x^2 - 3y^2)$. Thus, there are actually infinitely many 3-point quadrature formulas of strength 4, and we can choose (r, s) to ensure that any particular point is included. In this way, we see that there is no guarantee that the points $\{t_k\}$ need to be in the interval of integration, although practical people also want to choose points to minimize error. There will be an exercise or two on this, as well as one showing that there is exactly one choice of (r, s) for a strength 4 quadrature formula which actually has strength 5.