DIFFERENCE EQUATIONS AS BIOLOGICAL MODELS

Paul Cull

Received March 1, 2006

Abstract. Difference equations have been used in population biology from Fibonacci to May and Yorke and to the present day. They have also been used in other biological fields. Here, we describe a number of fairly well know examples. We give techniques and results about the analysis of linear and nonlinear difference equations. We also show that there are practical and theoretical limitations on the analysis of nonlinear models.

1 Introduction For almost 1000 years from Fibonacci's rabbits of 1200 to the present day, difference equations have been used as biological models. Here we will give a brief run through some of the rudiments of difference equations and describe some simple models used in biology. To keep the paper short, we will only discuss some models from population biology.

We start with the story of Fibonacci's rabbits, where we find and solve a simple difference equation. We then generalize to k^{th} order linear difference equations and outline the theory of these equations. We give stronger results about estimating and computing the solutions to nonnegative difference equations including those with inputs. We return to Fibonacci's model and show how it can be generalized to Leslie's matrix model. After a brief discussion of Leslie's model, we describe the generalization of these results contained in the Perron-Frobenius theorem. We then turn to a consideration of nonlinear models. We show that a number of these models can be analyzed quite easily, but that *chaos* and particularly *sensitive dependence on initial conditions* may make even simple appearing nonlinear models difficult to use in practice. We briefly show that *enveloping* can be used for many common population models. Finally, we consider *universality* and *undecidability* and argue that a complete theory of nonlinear difference equations is impossible.

1.1 Notation There are a few items of notation in this paper which may not be familiar. When we are talking about sequences, we often use the notation x_n to mean the n^{th} element of the sequence. But we also, ambiguously, use x_n to mean the entire sequence. When we want to be careful, we use the notation $\langle x_n \rangle$ to mean a whole sequence.

There are two notions of bounding with special notation. [16] We say that

$$f(n) \;=\; \Theta(\, g(n)\,)$$

iff there exist constants $c_1 > 0$ and $c_2 > 0$ so that there exists an n_0 and $\forall n \geq n_0$,

$$c_1 g(n) \leq \; f(n) \; \leq c_2 g(n).$$

We also say

$$f(n) \;=\; \mathcal{O}(\, g(n)\,)$$

2000 *Mathematics Subject Classification.* 92B05, 32A11, 39-02.

Key words and phrases. Difference equations, mathematical biology, history, population biology, chaos, stability, complexity, unsolvability.

if

$$f(n) \leq c_2 g(n),$$

that is if $g(n)$ is an upper bound but not necessarily a lower bound on $f(n)$.

2 Fibonacci's Rabbits In 1202 the Italian mathematician Fibonacci (also called Leonardo Pisano or Leonardo of Pisa) published *Liber Abaci*, a book of problems to illustrate the superiority of Arabic numerals over Roman numerals in arithmetic computation. (A translation by L. Sigler [25] has recently been published .) Fibonacci posed this seemingly easy problem about rabbits:

(a) You start at month 1 with one pair of immature rabbits.
(b) It takes rabbits 1 month to grow to reproductive maturity. After which, in each succeeding month each mature pair produces an immature pair.
(c) Rabbits live forever.
(d) **QUESTION:** After 13 months how many pairs of rabbits do you have?

To formalize this problem we can let M_n be the number of mature pairs in month n, and let I_n be the number of immature pairs in month n. Then by (b) and (c):

$$M_n = M_{n-1} + I_{n-1} \tag{1}$$

because each mature pair in M_{n-1} survives into month n and each immature pair in month $n-1$ becomes mature in month n. We also have:

$$I_n = M_{n-1}$$

because each mature pair in month $n-1$ produces an immature pair in month n. This problem can now be analyzed as a pair of coupled difference equations. But, it will convenient to use $I_{n-1} = M_{n-2}$ to re-write (1) as:

$$M_n = M_{n-1} + M_{n-2}. \tag{2}$$

To start calculating we need two initial initial values of M_n. We are given $I_0 = 1$ and $M_0 = 0$ and we can deduce that $M_1 = 1$. From this point there is no difficulty in computing M_{13} and M_{12}, and since $I_{13} = M_{12}$, the total number of rabbits will be $M_{13} + M_{12}$ which strangely enough is M_{14}.

We have a very neat difference equation, namely,

$$x_n = x_{n-1} + x_{n-2}. \tag{3}$$

which is satisfied by both M_n and I_n and even by their sum $M_n + I_n = T_n$. The three sequences $\langle M_n \rangle$, $\langle I_n \rangle$, and $\langle T_n \rangle$ do have different initial conditions, but these sequences are all simple shifts of the the Fibonacci sequence, $\langle f_n \rangle$,

$$0,\ 1,\ 1,\ 2,\ 3,\ 5,\ 8,\ \cdots$$

or if you'd rather the bi-infinite sequence

$$\cdots,\ -8,\ 5,\ -3,\ 2,\ -1,\ 1,\ 0,\ 1,\ 1,\ 2,\ 3,\ 5,\ 8,\ \cdots.$$

From the difference equation and the initial conditions it is no great feat to compute the 13^{th} or the 20^{th} term in these sequences. But, Fibonacci's point that this is easier done using Arabic rather than Roman numerals is well-taken.

For an understanding of the long term behavior of these sequences, we should step outside of the straight-jacket of the natural numbers. It is easy to show that

$$f_n \;=\; \frac{1}{\sqrt{5}}(\lambda_0^n \;-\; \lambda_1^n)$$

where $\lambda_0 \;=\; (1+\sqrt{5})/2$ and $\lambda_1 \;=\; (1-\sqrt{5})/2$. (E.g., one can show that this formula satisfies the initial conditions, and then verify the formula by induction.) This solution, sometimes called Binet's formula, was not available to Fibonacci because he did not have the notation to express it. (Modern algebraic notation is superior to Arabic numbers, as Arabic numbers are superior to Roman numerals.) Notice that this formula tells us a strange fact: we can take two *irrational numbers* raise them to a power, subtract them, and divide by an *irrational*, and the result always turns out to be a *natural* number. This is surprising because if we used other irrational numbers, we would get an irrational number rather than a natural number as the result.

This formula does display a fact that is not obvious from the direct computation. That is, the sequence grows exponentially with a factor of about 1.618 at each step. Also, since $|\lambda_1| \;<\; 1$ and $1/\sqrt{5} \;<\; 1/2$,

$$f_n \;=\; ROUND\Big(\frac{\lambda_0^n}{\sqrt{5}}\Big),$$

and the seemingly two-dimensional solution (depending on both λ_0 and λ_1) is really one-dimensional (depending on only λ_0). We will see that these features are **NOT** specific to the Fibonacci example, but they appear in solutions to a wide variety of difference equations.

Is the difference equation (3) a reasonable biological model? I believe that it is. It says that left alone a biological population should display exponential growth which seems reasonable. The model is obviously unreasonable in that it says that rabbits are immortal and eternally continue to reproduce. These features of the model are not serious objections if the model is only used to predict the population size for several generations. In making a model we should apply Occam's razor which says that among a variety of possibilities we should select the simplest. The only simpler model for the rabbits would be

$$x_n \;=\; \lambda_0\, x_{n-1}\;,$$

but here the simplicity of the difference equation comes at the cost of introducing an irrational multiplier. So, it seems to me that the Fibonacci difference equation **is** the simplest model.

We try to create **robust** models so that slight deviations from the model's assumptions do not change the model's predictions. As we will see the prediction of exponential population growth is a *robust* property shared by a variety of difference equation models.

3 Linear Difference Equations The Fibonacci equation can be generalized to the homogeneous linear equation with constant coefficients. Here, x_{n+k} is given as a *linear* function of $x_n, \ldots, x_{n+k-1}$. In other words, for all $n \geq 0$

$$x_n = c_1 x_{n-1} + \cdots + c_{k-1} x_{n-k+1} + c_k x_{n-k}\;.$$

where $c_1, \ldots, c_k$ are constants and $c_k \neq 0$. This is called a k^{th} order homogeneous linear difference equation with constant coefficients.

We will refer to the polynomial

$$ch(\lambda) = \lambda^k - c_1\lambda^{k-1} - \cdots - c_{k-1}\lambda - c_k$$

as the **characteristic polynomial** of the difference equation and its roots will be called the **eigenvalues**.

Theorem 1. *The solution to*

$$x_n = c_1 x_{n-1} + \cdots + c_{k-1} x_{n-k+1} + c_k x_{n-k}$$

is

$$x_n = \sum_{i=1}^{k} a_i \, \lambda_i^n = a_1 \, \lambda_1^n \; + \; \ldots \; + \; a_k \, \lambda_k^n$$

when $ch(\lambda)$ has k distinct roots, and, in general,

$$x_n = \sum_{i=1}^{j} p_i(n) \, \lambda_i^n = p_1(n) \, \lambda_1^n \; + \; \ldots \; + \; p_j(n) \, \lambda_j^n$$

where $\lambda_1, \ldots, \lambda_j$ are the distinct roots of $ch(\lambda)$ and $p_i(n)$ is a polynomial in n, and one plus the degree of $p_i(n)$ is the multiplicity of λ_i as a root of $ch(\lambda)$.

Asymptotic Behavior

1. ### Dominant Eigenvalue – Multiplicity 1

   If λ_1 is larger than every other eigenvalue and λ_1 is a simple root of $ch(\lambda)$ then

   $$\lim_{n \to \infty} \frac{x_n}{\lambda_1^n} \; = \; \text{constant.}$$

   The *constant* may be 0. In other words,

   $$|x_n| \; = \; \mathcal{O}(|\lambda_1|^n).$$

2. ### Dominant Eigenvalue – Multiplicity $d + 1$

   If λ_1 is larger than every other eigenvalue and λ_1 has multiplicity d+! as a root of $ch(\lambda)$ then

   $$\lim_{n \to \infty} \frac{x_n}{n^d \, \lambda_1^n} \; = \; \text{constant.}$$

   The *constant* may be 0. In other words

   $$|x_n| \; = \; \mathcal{O}(n^d \, |\lambda_1|^n).$$

3. ### Several Dominant Eigenvalues – Multiplicity 1

   If there are several simple eigenvalues, λ_1, λ_2, $\cdots$, λ_j all with largest magnitude then

   $$|x_n| \; = \; \mathcal{O}(|\lambda_1|^n).$$

   (Unlike in (2) limits may not exist.)

4. Several Dominant Eigenvalues – Several Multiplicities

If there are several eigenvalues, $\lambda_1, \lambda_2, \cdots, \lambda_j$ all with largest magnitude, and $d = \max\{d_1, d_2, \cdots, d_j\}$ where $d_i + 1$ is the multiplicity of λ_i as a root of $ch(\lambda)$,

$$|x_n| \;=\; \mathcal{O}(n^d\,|\lambda_1|^n).$$

(Again, limits may not exist.)

In each of these cases, we see that the k-dimensional solutions given in Theorem 1 asymptotically become 1-dimensional. So, some degree of long-range prediction of upper bounds for linear models is possible. If a few extra conditions are satisfied, then the limits can be shown to be non-zero and the $\mathcal{O}$'s can be replaced by Θ's, that is, the long range predictions become exact orders of growth rather than just upper bounds.

The theory here is quite general. The c_i's and the x_n's can be complex numbers. For reasonable biological models we may want to limit the coefficients to be real numbers, or nonnegative real numbers, or even natural numbers. We consider such restrictions in the following subsections.

3.1 Nonnegative Difference Equations In many biological situations (and also in other applications), we can assume that x_n really does behave like a quantity, that is, x_n is always nonnegative and usually positive. For these situations we can often model the process with a nonnegative difference equation. The important point is that analyses of these equations is easier. They have a dominant eigenvalue λ_0 which has multiplicity 1, and with the aperiodic condition, $\gcd\{i|c_i > 0\} = 1$, λ_0 is strictly bigger than every other eigenvalue. This allows us to get good bounds on the solutions to nonnegative homogeneous equations because, as we expect, the solutions will be very like λ_0^n. As we mentioned, this is again a reduction in dimension from k to 1.

We can also consider models which have an input. (Such models are called inhomogeneous or non-homogeneous.) Again, reasonable long-term predictions are possible. The general *rule-of-thumb* is that the larger of the input and λ_0^n (the solution to the corresponding homogeneous equation) should dominate the solution. There is even a reasonable prediction when the input has the form $p(n)\,\lambda_0^n$ in which $p(n)$ is a polynomial in n.

We summarize these long term prediction in the following outline.

1. Nonnegative Problem

(a) Difference Equation

$$x_n \;=\; c_1\,x_{n-1} \;+\; c_2\,x_{n-2} \;+\; \ldots \;+\; c_k\,x_{n-k} \;+\; g(n).$$

with $c_1 \geq 0$, $c_2 \geq 0$, $\ldots$, $c_{k-1} \geq 0$, and $c_k > 0$, and $g(n) \geq 0$.

(b) Positive Initial Conditions

$$x_1 \;>\; 0,\;\; x_2 \;>\; 0,\;\; \ldots\;\; x_k \;>\; 0.$$

(c) λ_0 is the unique positive real root of

$$\lambda^k \;-\; c_1\,\lambda^{k-1} \;-\; c_2\,\lambda^{k-2} \;-\; \ldots \;-\; c_k \;=\; 0.$$

2. Homogeneous ($g(n) \equiv 0$)

   (a) then $x_n = \Theta(\lambda_0^n)$.

3. Non-Homogeneous ($g(n) > 0$)

   (a) If $g(n) = \mathcal{O}(\gamma^n)$ with $\gamma < \lambda_0$
   then $x_n = \Theta(\lambda_0^n)$.

   (b) i. If $g(n) = \lambda_0^n$ then $x_n = \Theta(n \lambda_0^n)$.
   ii. If $g(n) = n^d \lambda_0^n$ then $x_n = \Theta(n^{d+1} \lambda_0^n)$.

   (c) If $g(n) = \Theta(\gamma^n)$ with $\gamma > \lambda_0$
   then $x_n = \Theta(\gamma^n)$.

3.2 Integer Solutions and Rounding As we observed before,

$$f_n = \text{Round}\left(\lambda_0^n/\sqrt{5}\right) \text{ for all } n \geq 0, \tag{4}$$

where Round(X) returns the integer nearest to X. Does such a result only hold for Fibonacci numbers? In particular, if one generalizes to the *k-binocci* numbers which satisfy $x_n = x_{n-1} + \cdots + x_{n-k}$, does such a rounding formula hold? The answer is YES. Two properties are used to prove rounding, nonnegativity of the difference equation, and bounding the absolute values of all eigenvalues (except λ_0) by 1. The following results are from [3].

Definition 3.1. *A difference equation $x_n = c_1 x_{n-1} + \cdots + c_k x_{n-k}$ is* **1-bounded** *iff*

$$\forall i \; c_i \in \mathbb{N}, \text{ and } c_k \in \mathbb{N}^+, \text{ and } \frac{\lambda - 1}{\lambda - \lambda_0} ch(\lambda) \text{ is a nonnegative polynomial,}$$

where $ch(\lambda) = \lambda^k - c_1\lambda^{k-1} - \cdots - c_k$ is the characteristic polynomial of the difference equation, and λ_0 is the unique positive root of $ch(\lambda)$. If, in addition, $\frac{\lambda-1}{\lambda-\lambda_0} ch(\lambda)$ is primitive (aperiodic), that is, $\gcd\{i | c_i > 0\} = 1$, the difference equation is **strongly 1-bounded**.

Theorem 2. *If x_n is an integer sequence which is a solution to a* **1-bounded** *difference equation, then there is an α so that*

a)
$$\forall n \geq 0 \qquad |x_n - \alpha\lambda_0^n| \leq \max_{0 \leq j \leq k-1} \{|x_j - \alpha\lambda_0^j|\}.$$

b) *If also*
$$\max_{0 \leq j \leq k-1} \{|x_j - \alpha\lambda_0^j|\} < 1/2$$
then $\forall n \geq 0$, $x_n = Round(\alpha\lambda_0^n)$.

c) *If the difference equation is* **strongly 1-bounded**
$$\exists n_0 \; \forall n \geq n_0 \qquad x_n = Round(\alpha\lambda_0^n).$$

The following theorem gives some easy to test *sufficient* conditions for rounding to hold. Since we're dealing with integer nonnegative polynomials, to test whether $c_1 + 1 \geq \lambda_0$ uses only one evaluation of the characteristic polynomial at an integer point

Theorem 3. *Assume x_n is an integer sequence which is a solution of the nonnegative difference equation $x_n = c_1 x_{n-1} + \cdots + c_k x_{n-k}$, so that $x_n = \alpha\lambda_0^n + d_n$ where λ_0 is the positive eigenvalue of the difference equation and d_n has no λ_0^n component. If*

$$c_{k-1} \geq \cdots \geq c_1 \text{ and } c_1 + 1 \geq \lambda_0$$

$$\text{and } \max\{|d_0|, |d_1|, \cdots, |d_{k-1}|\} < 1/2$$

then $x_n = Round(\alpha\lambda_0^n)$ for all $n \geq 0$.

If

$$c_{k-1} \geq \cdots \geq c_1 \text{ and } c_1 + 1 > \lambda_0$$

then there is an n_0 so that $x_n = Round(\alpha\lambda_0^n)$ for all $n \geq n_0$, and n_0 is the least integer so that $\max\{|d_{n_0}|, |d_{n_0+1}|, \cdots, |d_{n_0+k-1}|\} < 1/2$.

Here we have another example of dimension reduction. In this case, the solution is one-dimensional in spite of the k in the difference equation. In contrast to previous examples, here the reduction holds for all $n \geq 0$ rather than only holding asymptotically. To obtain these results we step out of the system in two ways – even though the sequences are integer sequences, we use a real (irrational) number λ_0; we also step out of the field of real or complex numbers by using *Round* which is a non-standard operation.

3.3 Computing Solutions By their very nature, difference equations give an algorithm for their solution:

> For a k^{th} order equation, use the k initial conditions to compute the next value. Treat the this new value and the last $k-1$ initial conditions as new initial conditions, and compute the next value. Repeat this procedure for computing a new value from the last k values.

Are there more efficient, i.e. quicker ways to compute solutions?

For example, the Fibonacci numbers can be computed as products of Lucas numbers. The Lucas numbers are the sequence that is a solution to the Fibonacci difference equation

$$l_n \;=\; l_{n-1} \;+\; l_{n-2}$$

with the initial conditions $l_0 \;=\; 2$ and $l_1 \;=\; 1$. This sequence satisfies the formula:

$$l_n \;=\; \lambda_0^n + \lambda_1^n \,.$$

It is relatively simple to show that if n is even then

$$l_{2n} \;=\; (l_n)^2 \;-\; 2.$$

So the Lucas numbers, for n a power of 2, can be computed quite quickly. In fact, about $\log n$ multiplies suffice. Since the numbers in this sequence grow quite quickly (l_n has $\Theta(n)$ digits), it would be reasonable to charge more for a multiplication of larger elements of the sequence. The classical multiplication algorithm use $\Theta(n^2)$ operations to multiply two n digit numbers. It's easy to see (using difference equations or summations) that in the series of multiplications to compute a large Lucas number, the last multiply is really the only one that counts. The rest of the multiplies all together cost at most 1/3 of the final multiply.

Computing these Lucas numbers also gives a method for computing Fibonacci numbers. It is easily seen by multiplying the two "Binet" formulas that

$$f_{2n} \;=\; f_n\, l_n.$$

So, for n a power of 2, the n^{th} Fibonacci number can be computed as a product of Lucas numbers which can be iteratively computed by squaring and subtracting 2. With classical multiplication, the n^{th} Fibonacci number and the n^{th} Lucas number can be computed in $\Theta(n^2)$ time. But as shown in Knuth [17], there are faster multiplication algorithms, and even a sequence of algorithms so that for every $\epsilon > 0$ there is an a multiplication algorithm which runs in $\Theta(n^{1+\epsilon})$. Hence using these faster multiplication techniques, the Fibonacci numbers can be computed more quickly than by using the obvious addition algorithm. The addition method is a $\Theta(n^2)$, and the faster methods are almost $\Theta(n)$, so we expect the faster methods to be much faster. Unfortunately, the faster multiplication methods are only faster for large values of n. The addition method may be faster for small n.

We discuss various algorithms for Fibonacci numbers and give some timing data on run times of implementations of these methods in our paper [9].

4 Multi-dimensional Difference Equations

4.1 Leslie's Model In the Fibonacci story the rabbits are immortal, but with a few exceptions, like the Energizer Bunny, real rabbits have small finite lifetimes. In matrix form, this model is:

$$\begin{pmatrix} M_{t+1} \\ I_{t+1} \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} M_t \\ I_t \end{pmatrix}.$$

where the 1 in the upper left of the matrix represents the immortality assumption. If we flip M and I, we get

$$\begin{pmatrix} I_{t+1} \\ M_{t+1} \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{pmatrix} I_t \\ M_t \end{pmatrix},$$

which has the same form as the previous model, but we can change the interpretation to make a more biologically reasonable model. Here we can let M_t mean the number of new pairs and let I_t be the number of old pairs. The two 1's in the first row now say that each pair (new or old) produces a new pair at every time step. The 1 in the lower left indicates that each new pair becomes an old pair. The 0 in the lower right means that each old pair then dies. To make this model even more reasonable, we could introduce s_1 to be the probability of surviving from the new to the old group. We could also assume that the number of offspring produced depends on the age class. With these assumptions, the model becomes

$$\begin{pmatrix} I_{t+1} \\ M_{t+1} \end{pmatrix} = \begin{bmatrix} f_1 & f_2 \\ s_1 & 0 \end{bmatrix} \begin{pmatrix} I_t \\ M_t \end{pmatrix}.$$

We can generalize the Fibonacci model with two age classes to a model with k age classes. In population biology the model with k age classes is usually called **Leslie's model**. In 1945, Leslie [18] published one of the most influential papers in population biology. In it he introduced a generation of biologists to vectors and matrices. The model Leslie described is quite similar to the renewal model [14] which was already used in population biology, see for example, Lotka's 1925 book, *Elements of Mathematical Biology* [20].

The Leslie model can be concisely stated as

$$X_{t+1} = LX_t$$

where X_t and X_{t+1} are population vectors and L is a Leslie matrix. A **Leslie matrix** contains both survival rates and fertility rates, specifically,

$$L = \begin{bmatrix} f_1 & f_2 & \dots & \dots & f_k \\ s_1 & 0 & \dots & \dots & 0 \\ & s_2 & & & \vdots \\ & & \ddots & & \vdots \\ & & & s_{k-1} & 0 \end{bmatrix}.$$

The first row of L consists of fertility rates where f_i is the number of offspring (newborn) produced by an individual of age class i in one time unit, and the subdiagonal of L contains the survival rates where s_i is the probability that an individual in age class i will survive to age class $i+1$. All other entries in L are zero.

The usual assumptions on these parameters are that for each i, $0 < s_i \leq 1$ and $f_i \geq 0$. The first assumption makes sense if one interprets s_i as a probability and assumes that there is some possibility for an individual to survive from a particular age class into the next age class. Further, if any s_i were zero, then in $k-i$ steps the population would become a population in which the last $k-i$ age classes are empty, and all future developments occur within and depend only on the first i age classes. For similar reasons, one usually assumes that $f_k \geq 0$. That is, if one or several of the oldest age classes have zero fertility, then the composition of these older age classes has no effect on the rest of the population and in a small number of steps the composition of these age classes is determined by the younger age classes with no effect from the original composition of these oldest age classes.

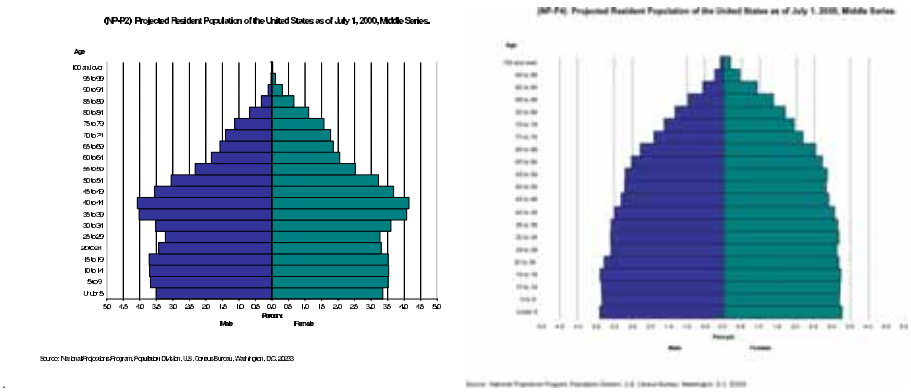
An extra assumption made in Leslie's original model and often used in demographic applications is that at least two adjacent fertility rates are positive. This assumption is often enforced by averaging fertilities. That is, the number of offspring from females in each age class is measured, but a fraction of these are attributed to females in the next age class because the females are assumed to be aging as the measurements are taken. A mathematically more appropriate assumption, which includes the Leslie assumption as a special case, is that there is a power of the Leslie matrix which is strictly positive. Luckily this can be checked easily using the greatest common divisor of the indices of positive fertility rates.

Theorem 4. *Let L be a Leslie matrix. Then there exists an $m \geq 0$ with $L^m \gg 0$ iff $\gcd\{i | f_i > 0\} = 1$, where $A \gg 0$ means that every entry in the matrix A is strictly positive.*

This convergence of the Leslie matrix is usually not the result that is used. Instead, biologists look for the **stable age distribution**, and see how the population is converging to this distribution. The **stable age distribution**, D, is the (unique, positive) eigenvector associated with the positive eigenvalue λ_0. In the aperiodic case (forced by the Leslie's averaging of fertility rates), for every non-negative initial vector X_0 the population vector converges to a multiple of D, in the sense that

$$\lim_{n\to\infty} \frac{X_n}{\lambda_0^n} = \alpha D.$$

The scalar constant does depend on the initial X_0. Under the further assumption that $\lambda_0 > 1$, The components of D decrease in a roughly exponential fashion, that is, $D[i] \leq D[i-1]/\lambda_0$. From its decreasing shape the stable age distribution is sometimes called the *inverted pyramid* distribution.



(a) Projected population distribution for 2003.

(b) Projected population distribution for 2050.

Figure 1: Two population predictions from the Bureau of the Census.

The graphs in Figure 1 show the census bureau's estimates for the US population in 2003 and 2050. Both demonstrate a roughly "pyramidal" rather that the "inverted pyramid" of the Leslie model because these graphs have the older age classes at the top rather than at the bottom. The graphs also show the "pig in a python" caused by the "baby boom". As time goes on the wide part of the pyramid passes through all the age classes and dies out.

A weaker form of convergence occurs even if the aperiodic condition is not fulfilled. Let $x_i(n)$ be the i^{th} component of X_n, then simply assuming that X_0 has all positive components (or that there is some m so that X_0 has all positive components) implies that for each i,

$$x_i(n) \;=\; \Theta(\lambda_0^n).$$

But here the distribution is not necessarily the stable age distribution. More details about the periodic case are in [10] [11] [8]. In spite of Leslie's desire to avoid the periodic model, the use of such a model in population biology was discussed by Bernardelli [2] a few years before Leslie's paper.

4.2 Perron-Frobenius Theory The Perron-Frobenius Theorem simplifies working with nonnegative matrix difference equations. Provided that the matrix is *primitive*, the equations are as simple as the Leslie models and not much more complicated than scalar equations.

Definition 4.1. *A matrix M is* **nonnegative** *if every element $m_{i,j}$ is ≥ 0. A matrix M is* **primitive** *if M is nonnegative and there is a positive integer t so that every element in M^t is strictly positive, which we will write as $M^t \gg 0$.*

Theorem 5 (Perron-Frobenius). *If M is a primitive matrix then*

(a) M has a maximum positive real eigenvalue λ_0
(b) λ_0 has multiplicity 1, (it is a simple root of the characteristic polynomial)
(c) for every other eigenvalue λ_i, $\lambda_0 > |\lambda_i|$ (it is strictly dominant)
(d) $\min_i \sum_j m_{i,j} < \lambda_0 < \max_i \sum_j m_{i,j}$
$\min_j \sum_i m_{i,j} < \lambda_0 < \max_j \sum_i m_{i,j}$
(e) the row and column eigenvectors associated with λ_0 are strictly positive, and are unique up to scaling

(f) the sequence M^t is asymptotically one-dimensional, its columns converge to the column eigenvector associated with λ_0, and its rows converge to the row eigenvector associated with λ_0.

We will not prove this theorem; proofs appear in many places including [1, Chapter 2] and [24, Chapter 1].

The Perron-Frobenius result is somewhat weaker than the Leslie results. In particular the *inverted pyramid* is not promised for the stable distribution. On the other hand, the Perron-Frobenius applies to a much wider range of models including, for example, finite Markov chain models. *Graph theory* is very useful in working with these models. It often gives a nice pictorial description, and can even lead to improved computational methods. For example, using graph theory one can show that determining if an $n \times n$ matrix is *primitive* can be accomplished in $\Theta(n^2)$ time which is much faster than the time for even one matrix multiplication.

5 Nonlinear Equations Let us now consider one-dimensional equations of the form

$$x_{n+1} = f(x_n).$$

There is essentially only one linear one-dimensional equation, but to paraphrase Tolstoy: Equations can be linear in only one way, but equations can be nonlinear in many different ways. So we should not expect to have a general theory for nonlinear equations. Rather, we hope to have different theories for different classes of nonlinear equations.

As a simple example, consider

$$x_{n+1} = \sqrt{x_n}\,.$$

For this equation to make sense, we assume that $x_0 \geq 0$ and that $\sqrt{x}$ returns the nonnegative square root of x. We can calculate some iterates

$$\begin{aligned} x_1 &= x_0^{1/2} \\ x_2 &= x_1^{1/2} = (x_0^{1/2})^{1/2} = x_0^{1/4} \\ x_3 &= x_2^{1/2} = (x_0^{1/4})^{1/2} = x_0^{1/8} \end{aligned}$$

and see that the solution is

$$x_n = x_0^{1/2^n}.$$

There are four cases:

(a) $x_0 = 0$, and then $x_n = 0$ for all $n \geq 0$,
(b) $0 < x_0 < 1$, and then $1 > x_{n+1} > x_n > 0$,
(c) $x_0 > 1$, and then $1 < x_{n+1} < x_n$,
(d) $x_0 = 1$, and then $x_n = 1$ for all $n \geq 0$.

We summarize these cases by saying that 0 and 1 are **fixed points** of the system since $f(p) = p$ for $p = 0, 1$. The fixed point 0 is **unstable**, while 1 is a **stable fixed point** which attracts all solutions with $x_0 > 0$.

Some nonlinear equations have **oscillations**. For example,

$$x_{t+1} = \frac{1}{x_t}\,.$$

with initial condition x_0 will generate the sequence $x_0, 1/x_0, x_0, 1/x_0, \cdots$ and we say that this sequence is an oscillation of **period 2**.

Fixed points can have *local* instead of *global* stability, which we'll demonstrate in the next example. For a differentiable function, $f(x)$, local stability at the fixed point, p, requires

$$|f'(p)| \leq 1.$$

In general, this condition does not imply global stability, in the next subsection *Enveloping* we will see that for many of the usual population models, local stability does imply global stability.

The difference equation $x_{t+1} = f(x_t)$, with the following:

$$f(x) = \begin{cases} 3x & 0 \leq x < 1/2 \\ (9-6x)/4 & 1/2 \leq x \leq 3/4 \\ (6-2x)/4 & 3/4 \leq x \leq 5/4 \\ (11-6x)/4 & 5/4 \leq x \leq 11/6 \\ 0 & x \geq 11/6 \end{cases}$$

displays a several features common in population models. Here, the $3x$ for $x \in (0, 1/2)$ indicates the expected exponential growth when the population size is small – less than $1/2$ of the equilibrium value which is normalized to 1. If the population size is very large – at least $11/6$ of the equilibrium value, then the population crashes to 0. In between these two extremes, there are some interesting phenomena. Near $x = 1$, $f(x)$ is a straight line with slope $-1/2$. Since $-1 < f'(x) < 0$ and $f(1) = 1$, $x = 1$ is a *locally stable fixed point* and trajectories in this region should display damped oscillations of period 2. For example, one of these trajectories is

$$\frac{3}{4}, \quad \frac{9}{8}, \quad \frac{15}{16}, \quad \frac{33}{32}, \quad \frac{63}{64}, \quad \cdots$$

and as you can see these iterates are rapidly approaching 1. There is also a period 2 cycle, i.e.,

$$\frac{1}{2} \quad \longrightarrow \quad \frac{3}{2} \quad \longrightarrow \quad \frac{1}{2} \quad \longrightarrow \quad \cdots.$$

Although it is possible to give a full analysis of all starting points for this model we shall not do so here. The main point about this model is that it is similar to various population models and that it does display *local* stability without *global* stability, e.g. a trajectory starting at $1/2$ will stay in a cycle and **not** converge to 1.

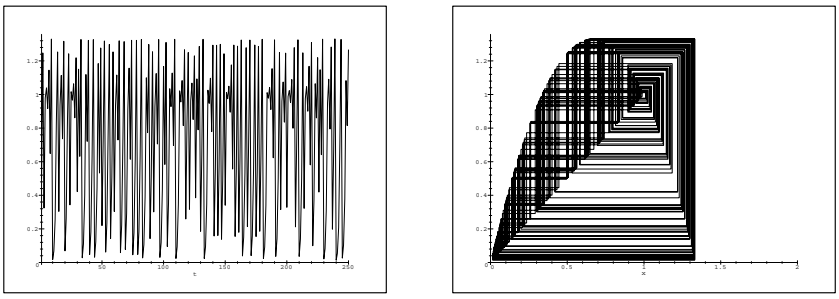
Our point, so far, is that *some* nonlinear equations are not difficult to analyze. But there are nonlinear equations which are more difficult.

As Yorke [19] and May [21, 22] have shown even seemingly simple models, such as,

$$x_{n+1} = x_n\,(1 + r\,(x_n - 1)).$$

are exceedingly hard to understand because of a phenomena called **chaos**. (See Figure 2.) Basically, *chaos* means that it is very difficult to predict long term behavior from the model and measured initial conditions. In this model r is the reproductive rate. This model is predictable when r is small, but chaos occurs for larger values of r.

Specifically, these models display the *butterfly effect* which is also called **sensitive dependence on initial conditions**. The metaphor is that the flapping or non-flapping of a butterfly's wings in Borneo can affect the rainfall in Brazil. Less picturesquely, to answer a question like is $x_n > 1$? or is $x_n > 1/2$? requires more and more knowledge of x_0 as n increases. So, we might be able to predict the value of x_1 if we know x_0 to a few bits of accuracy, but to predict x_{100} we would need a much more accurate estimate of x_0. This sensitive dependence puts a *practical* limitation on our use of such models. In practice, we can usually measure data to an an accuracy of a few digits, but (at least when r is large) long term predictions from *chaotic* models will be practically meaningless.



(a) Chaos (x_n as a function of n.) (b) Web plot (x_{n+1} as a function of x_n.)

Figure 2: A chaotic trajectory and its web plot; quadratic map with $r = 2.99$.

5.1 Enveloping Contrary to the chaos in the above section, stability properties of many nonlinear population models can be dealt with a a technique that we call **enveloping**.[5] The idea is that the function $f(x)$ in the difference equation is bounded by a self-inverse function. The bounding is from above for x's less than the fixed point of $f(x)$ and the bounding is from below for x's greater than the fixed point of $f(x)$. Under these circumstances *the* fixed point is *globally stable.* The surprising result is that for the commonly used population models this bounding function can be chosen to be a simple ratio of two linear functions. Here, we give some of the results and refer the reader to our other papers for more details. [7, 6, 12, 4]

A ***linear fractional function*** is a function of the form

$$\phi(x) = \frac{1 - \alpha x}{\alpha - (2\alpha - 1)x} \qquad \text{where } \alpha \in [0, 1) \ .$$

These functions have the properties

- $\phi(1) = 1$
- $\phi'(1) = -1$
- $\phi(\phi(x)) = x$
- $\phi'(x) < 0$.

Theorem 6. *Let $\phi(x)$ be a monotone decreasing function which is positive on $(0, x_-)$ and so that $\phi(\phi(x)) = x$. Assume that $f(x)$ is a continuous function such that:*

- $\phi(x) > f(x)$ *on* $(0, 1)$
- $\phi(x) < f(x)$ *on* $(1, x_-)$
- $f(x) > x$ *on* $(0, 1)$
- $f(x) < x$ *on* $(1, \infty)$
- $f(x) > 0$ *on* $(1, x_\infty)$

then for all $x \in (0, x_\infty)$, $\lim_{k\to\infty} f^{(k)}(x) = 1$.

The following seven fairly standard population models can be shown to be globally stable by enveloping with a linear fractional:

- $x_{t+1} = x_t e^{r(1-x_t)}$
- $x_{t+1} = x_t[1 + r(1 - x_t)]$
- $x_{t+1} = x_t[1 - r \ln x_t]$
- $x_{t+1} = x_t(\frac{1}{b+cx_t} - d)$
- $x_{t+1} = \frac{(1+ae^b)x_t}{1+ae^{bx_t}}$
- $x_{t+1} = \frac{(1+a)^b x_t}{(1+ax_t)^b}$ with $a > 0$, $b > 0$
- $x_{t+1} = \frac{rx_t}{1+(r-1)x_t^c}$.

To understand what we are doing take a look at Figure 3. In the first panel, we've plotted $f(x)$ from the sixth model in the above list, namely,

$$x_{t+1} = \frac{(1+a)^b x_t}{(1+ax_t)^b} = f(x_t)$$

using $b = 2$ and two different values of a. These two curves show how $f(x)$ sharpens as a is increased. The *dotted* line shows $1/x$. As you can see, $1/x > f(x)$ on the open interval $(0, 1)$, and $1/x < f(x)$ on the open interval $(1, \infty)$. This *enveloping* demonstrates that for $b = 2$, the fixed point 1 is globally stable. In fact, for $b \leq 2$ and $a > 0$, the linear fractional $1/x$ serves as an enveloping curve and establishes global stability.

When $b > 2$, the situation is more complicated. If the local stability condition, $ab \leq 2\,(1 + a)$ is satisfied, then it can be shown that a linear fractional envelopes $f(x)$, and so $x = 1$ is the globally fixed point, but the linear fractional used does depend on the values of the parameters a and b. In the second panel of Figure 3, we use $b = 3$ and $a = 2$ giving

$$x_{t+1} = \frac{(1+a)^b x_t}{(1+ax_t)^b} = \frac{27x_t}{(1+2x_t)^3}$$

which is represented as the solid curve. For an enveloping function we use the linear fractional with $\alpha = 1/4$ giving

$$\phi(x) = \frac{4 \; - \; x}{1 \; + \; 2\,x}$$

which is represented by the dotted curve.

6 Universality Surprisingly enough, difference equations are completely general. They capture the ideal of general computability. According to Turing's analysis [26], the state of any computation can be represented by a natural number, say x. This state uniquely determines the next state of the computation by following the Turing transformation $\mathfrak{T}$. So, *any* computation can be represented in the form

$$x_n \; = \; \mathfrak{T}(x_{n-1})$$

where $\mathfrak{T}$ represents the *program* for the computation. But, Turing also argued that there was a *universal* program (the Universal Turing Machine) which could carry out *any* computation.[15]

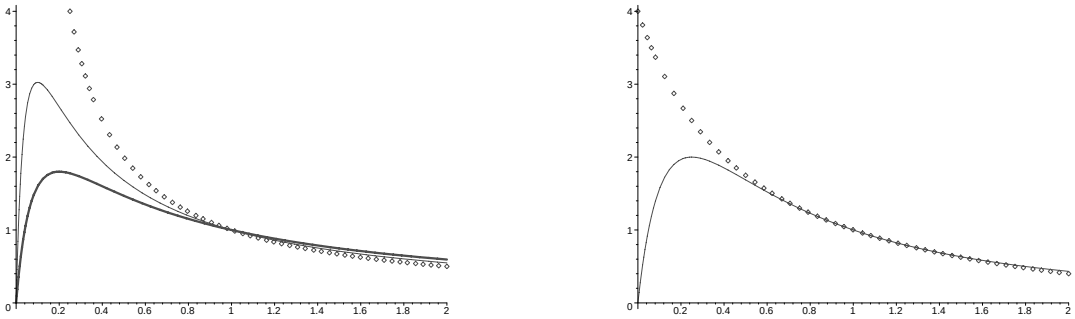


Figure 3: Two examples of model 6. The model with $b \leq 2$ is enveloped by $1/x$. (The curve becomes steeper as a is increased.) With $b = 3$ and $a = 2$ the model $27x/(1+2x)^3$ is enveloped by $(4-x)/(1+2x)$.

This universal machine, let's call it M, takes two inputs, a state x and a program p. It then computes the next state by

$$x_n \;=\; \mathfrak{T}_M(x_{n-1},\, p)$$

where $\mathfrak{T}_M$ represents the operation of this universal machine M. Then clearly

$$(x_n,\, p) \;=\; (\,\mathfrak{T}_M(x_{n-1},\, p),\, p).$$

We can reduce this pair of natural numbers to a single natural by using an invertible **pairing function**, $\mathbb{P}$. For example, we can use

$$\mathbb{P}(x,y) \;=\; \frac{(x+y)(x+y+1)}{2} \;+\; y$$

as a pairing function because it maps pairs of naturals one-to-one onto the naturals. Letting $y_n \;=\; \mathbb{P}(x_n,\, p)$ we get

$$\begin{aligned}
y_n \;&=\; \mathbb{P}(\,\mathfrak{T}_M(x_{n-1},\, p),\, p) \\
&=\; \mathbb{P}(\,\mathfrak{T}_M(\mathbb{P}_1^{-1}(\mathbb{P}(\,x_{n-1},\, p)),\quad \mathbb{P}_2^{-1}(\mathbb{P}(\,x_{n-1},\, p)),\quad p) \\
&=\; \mathbb{P}(\,\mathfrak{T}_M(\mathbb{P}_1^{-1}(\,y_{n-1}\,),\quad \mathbb{P}_2^{-1}(\,y_{n-1}\,)\,),\quad \mathbb{P}_2^{-1}(\,y_{n-1}\,)\,) \\
&=\; \mathfrak{U}(\,y_{n-1}\,)
\end{aligned}$$

where we've invented the function $\mathfrak{U}(\,y_{n-1}\,)$ to emphasize that the right hand side is solely a function of y_{n-1}. This function $\mathfrak{U}()$ is a **universal** function so that the **universal** difference equation

$$y_n \;=\; \mathfrak{U}(\,y_{n-1}\,)$$

contains within itself every possible computation. This means, in particular, that given $\mathfrak{U}()$ and y_0, we cannot, in general, decide if y_n goes to a fixed point or eventually cycles or ever visits a particular value. All of these questions are equivalent to the infamous **Halting Problem** and Turing showed that **NO ALGORITHM CAN SOLVE THE HALTING PROBLEM**.(see [13])

One might object that this is only true *in theory* because the function $\mathfrak{U}$ must be so complicated that it will never arise in practice. On the contrary, $\mathfrak{U}$ can be computed as

follows:

$$\begin{aligned}(L,\, R,\, q) &:= \text{UNPACK}(x)\\ \alpha &:= h(q,\ L \bmod 2)\\ \beta &:= g(q,\ L \bmod 2)\\ \hat{L} &:= \alpha\, L/2 \ +\ (1-\alpha)(2L\ +\ \beta)\\ \hat{R} &:= \alpha\,(2R\ +\ \beta)\ +\ (1-\alpha)\ R/2\\ \hat{q} &:= \delta(q,\ L \bmod 2)\\ \hat{x} &:= \text{PACK}(\hat{L},\, \hat{R},\, \hat{q}).\end{aligned}$$

Here, PACK($\hat{L}$, $\hat{R}$, $\hat{q}$) is an invertible function which maps triples of naturals to a single natural. (For example, we could use PACK($\hat{L}$, $\hat{R}$, $\hat{q}$) $=$ $\mathbb{P}(\mathbb{P}(L,\, R),\, q)$.) Both h and g are Boolean valued functions, so $\alpha \in \{0, 1\}$ and $\beta \in \{0, 1\}$. Further, δ is a function from $Q \times \{0, 1\}$ to Q, where Q is a finite set. The set Q can be reasonably small. A few dozen elements are enough. For the details of this construction see Minsky. [23]

Here one difference equation

$$y_n \ =\ \mathfrak{U}(\, y_{n-1}\,)$$

contains all *computable* sequences. For each *computable* sequence $\langle x_k \rangle$ there is a Turing machine program $p_{\langle x \rangle}$ so that when $p_{\langle x \rangle}$ is given k as an input, $p_{\langle x \rangle}$ will eventually halt and output the value of x_k. {At the risk of a slight confusion, we're using $p_{\langle x \rangle}$ for a Turing machine, for a program, and for the natural number representing this program.} so that Setting $L_0 = \mathbb{P}(p_{\langle x \rangle},\, k)$ and starting at $y_0 = \text{PACK}(L_0,\, 0,\, q_0)$ the sequence $\langle y_n \rangle$ reaches a fixed point which when UNPACKed yields the triple $(x_k,\, 0,\, q_{STOP})$.

7 Conclusions In this brief sketch, we can only outline some of the theory and applications of difference equations. More details can be found in our book *Difference Equations*. [8]

Our big points are:

(a) difference equations are useful biological models
(b) linear models have a very simple theory
(c) nonnegativity often implies stronger conclusions
(d) stepping outside a system may simplify the analysis of the system
(e) some nonlinear models can be analyzed
(f) *chaos* puts a practical limitation on what can be predicted from data
(g) *unsolvability* implies that a complete theory of nonlinear difference equations is impossible.

References

[1] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. SIAM, Philadelphia, PA, 1994.

[2] H. Bernardelli. Population waves. *J. Burma Res. Soc.*, 31:3–18, 1941.

[3] R. M. Capocelli and P. Cull. Rounding the solutions of Fibonacci-like difference equations. *Fibonacci Quarterly*, 41:133–141, 2003.

[4] P. Cull. Convergence of iterations. In R. Moreno-Diaz et al, editor, *EUROCAST 2005. LNCS 3643*, pages 457–466. Springer-Verlag, Berlin Heidelberg, 2005.

[5] P. Cull. Enveloping Implies Global Stability. In L. Allen B., Aulbach, S. Elaydi, and R. Sacker, editors, *Difference Equations and Discrete Dynamical Systems*, pages 170–181. World Scientific, Hackensack, NJ, 2005.

[6] P. Cull and J. Chaffee. Stability in discrete population models. In D. M. Dubois, editor, *Computing Anticipatory Systems: CASYS'99*, pages 263–275. Conference Proceedings 517, American Institute of Physics, Woodbury, NY, 2000.

[7] P. Cull and J. Chaffee. Stability in simple population models. In *Cybernetics and Systems 2000*, pages 289–294. Austrian Society for Cybernetics Studies, 2000.

[8] P. Cull, M. Flahive, and R. Robson. *Difference Equations: From Rabbits to Chaos*. Springer, New York, 2005.

[9] P. Cull and J. Holloway. Computing Fibonacci numbers quickly. *Information Processing Letters*, 32:143–149, 1989.

[10] P. Cull and A. Vogt. Mathematical analysis of the asymptotic behavior of the Leslie population matrix model. *Bulletin of Mathematical Biology*, 35:645–661, 1973.

[11] P. Cull and A. Vogt. The periodic limit for the Leslie model. *Mathematical Biosciences*, 21:39–54, 1974.

[12] Paul Cull. Stability in One-dimensional Models. *Scientiae Mathematicae Japonicae*, 58:349–357, 2003.

[13] M. Davis. *Computability and Unsolvability*. Dover, New York, NY, 1982.

[14] W. Feller. *An Introduction to Probability Theory and its Applications*. John Wiley, New York City, NY, 1968.

[15] R. Herken, editor. *The Universal Turing Machine*. Oxford University Press, Oxford, UK, 1988.

[16] D. Knuth. Big Omicron and Big Omega and Big Theta. *SIGACT News*, 8:18–24, April-June 1976.

[17] D. E. Knuth. *The Art of Computer Programming*. Addison-Wesley, New York City, NY, third edition, 1997.

[18] P. H. Leslie. On the use of matrices in certain population mathematics. *Biometrika*, 33:183–212, 1945.

[19] T-Y. Li and J. Yorke. Period three implies chaos. *American Mathematical Monthly*, 82:985–992, 1975.

[20] A. J. Lotka. *Elements of Mathematical Biology*. Dover Publications, New York City, NY, 1956.

[21] R. M. May. Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos. *Science*, 186:645–647, 1974.

[22] R. M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261:459–467, 1976.

[23] M. Minsky. *Computation: Finite and Infinite Machines*. Prentice-Hall, Englewood Cliffs, NJ, 1967.

[24] E. Seneta. *Non-negative Matrices*. John Wiley & Sons, New York City, NY, 1973.

[25] L. Sigler. *Fibonacci's Liber Abaci*. Springer-Verlag, New York City, NY, 2002.

[26] A. Turing. On Computable Numbers, with an Application to the Entscheidungsproblem. In M. Davis, editor, *The Undecidable*, pages 116–154. Raven Press, Hewlett, NY, 1965.

Paul Cull
Computer Science Dept., Oregon State University
Corvallis, OR 97331 USA
Email: pc@eecs.oregonstate.edu