Linear Recurrence Relations: The Theory Behind Them

David Soukup

October 29, 2019

Contents

1 Foreword

This guide is intended mostly for students in Math 61 who are looking for a more theoretical background to the solving of linear recurrence relations. A typical problem encountered is the following: suppose we have a sequence defined by

$$a_n = 2a_{n-1} + 3a_{n-2} \qquad \text{where } a_0 = 0, a_1 = 8.$$

Certainly this recurrence defines the sequence $\{a_n\}$ unambiguously (at least for positive integers n), and we can compute the first several terms without much problem:

$$\begin{aligned}
a_0 &= 0 \\
a_1 &= 8 \\
a_2 &= 2 \cdot a_1 + 3 \cdot a_0 = 2 \cdot 8 + 3 \cdot 0 = 16 \\
a_3 &= 2 \cdot a_2 + 3 \cdot a_1 = 2 \cdot 16 + 3 \cdot 8 = 56 \\
a_4 &= 2 \cdot a_3 + 3 \cdot a_2 = 2 \cdot 56 + 3 \cdot 16 = 160 \\
&\vdots
\end{aligned}$$

While this does allow us to compute the value of a_n for small n, this is unsatisfying for several reasons. For one, computing the value of a_{100} by this method would be time-consuming. If we solely use the recurrence, we would have to find the previous 99 values of the sequence first. Moreover, until we compute it we have no good idea of *how large* a_{100} will be. We can observe from the table above that a_n grows quite rapidly as n increases, but we do not know how to extrapolate this rate of growth to all n.

Most importantly, we would like to know the *order of growth* of this function. This is particularly relevant in computer science, where a solution for the amount of time a computer program will have to run, or the number of cases it will have to check, is frequently given by recurrence relations such as the above. Typically, we are not concerned so much with the values of the sequence at any particular n, but with how our values scale as n increases. This can be the difference between an algorithm that is viable and one that is useless for large data sets.

To begin, we recall the 'standard' way of solving these relations in Math 61. Since we have a linear recurrence, we can construct the *characteristic polynomial* associated to it:

$$t^2 - 2t - 3 \tag{1}$$

We find the roots by factoring this polynomial to get $(t-3)(t+1)$, so the roots are -1 and 3. So we make the assumption that our solution is of the form

$$a_n = c_1(-1)^n + c_2(3)^n$$

To find the constants, we use our remaining pieces of information, namely the initial conditions:

$$0 = a_0 = c_1(-1)^0 + c_2(3)^0$$
$$8 = a_1 = c_1(-1)^1 + c_2(3)^1$$

which gives us the system of equations

$$0 = c_1 + c_2$$
$$8 = -c_1 + 3c_2$$

Solving this system[1] gives the solution $c_1 = 2$, $c_2 = -2$. So we have our solution, namely

$$a_n = 2 \cdot 3^n - 2 \cdot (-1)^n$$

We can check that this agrees with our calculated values above; for instance, $a_4 = 2 \cdot 3^4 - 2 \cdot (-1)^4 = 2 \cdot 81 - 2 = 160$ as above. We mentioned the idea of computing a_{100}, but the formula tells us that $a_{100} \approx 2 \cdot 3^{100} \approx 1.03 \cdot 10^{48}$, a very large number[2].

Formally speaking, we have not proved this formula yet, though we could do it fairly easily, through induction on n. In setting the coefficients, we made sure the cases $n = 0$ and $n = 1$ worked. So, suppose our formula is true for $k - 1$ and $k - 2$. Then

$$\begin{aligned} a_k &= 2a_{k-1} + 3 \cdot a_{k-2} \\ &= 2 \cdot (2 \cdot 3^{k-1} - 2 \cdot (-1)^{k-1}) + 3 \cdot (2 \cdot 3^{k-2} - 2 \cdot (-1)^{k-2}) \\ &= (4 + 2) \cdot 3^{k-1} + (-4 + 6)(-1)^{k-1} \\ &= 2 \cdot 3^k - 2 \cdot (-1)^k \end{aligned}$$

as required. Note that since we are using the previous two cases in our induction, we needed to have two base cases to make it work.

The preceding was a valid mathematical proof of our general formula for a_n, but it rings hollow. Perhaps we can guess that a_n is growing exponentially, but why should this mysterious characteristic polynomial govern what types of powers we are allowed to use? In this guide, we will discover two methods which allow us to derive the solution without such chicanery. By no means are these the only possible approaches, but they are remarkable both in their ingenuity and how they use ideas developed in lower-division math classes in different ways. The first relies on the linear algebra ideas introduced in Math 33A: eigenvalues and the diagonalization of a matrix. The second requires only some familiarity with the idea of a power series, and some of the basic ideas in Math 31B. And either can be adapted to solve more complicated recurrences.

[1] For instance, we could add the two equations to obtain $8 = 4c_2$, meaning $c_2 = 2$. Then the first equation tells us $c_1 = -2$.

[2] Specifically, $a_{100} = 2 \cdot 3^{100} - 2 =$

1 030 755 041 464 022 662 072 922 259 531 242 545 404 215 044 000.

2 The matrix diagonalization method

(Note: For this method we assume basic familiarity with the topics of Math 33A: matrices, eigenvalues, and diagonalization.)

We return to our original recurrence relation:

$$a_n = 2a_{n-1} + 3a_{n-2} \qquad \text{where } a_0 = 0, a_1 = 8. \tag{2}$$

Suppose we had a computer calculate the 100th term by the direct computation method at the beginning of the foreword. Rather than storing the values of a_n for all n up to and including 100, note that at every step we do not use all the previous terms in the sequence. In particular, we need only save two pieces of information - the two previous terms - at every step. We can then compute the next pair of points as follows:

$$(a_n, a_{n-1}) = (2a_{n-1} + 3a_{n-2}, a_{n-1})$$

That is, we find the next term in the sequence, and we have to keep a_{n-1} around for the next iteration. This looks like a simple rewriting, but notice that we now have the vector (a_n, a_{n-1}) written as a *linear transformation* of the previous one. In matrix form, this looks like

$$\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ a_{n-2} \end{bmatrix}.$$

Since we now can go from one pair of terms to the next with just a matrix muliplication, we can step forward n terms just by muliplying the matrix $n-1$ times:

$$\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix}^{n-1} \begin{bmatrix} a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix}^{n-1} \begin{bmatrix} 8 \\ 0 \end{bmatrix}. \tag{3}$$

Now we've translated the problem into a mechanical one about matrix multiplications. We might say that we're done at this stage; but computing high powers of a matrix requires a lot of work in general[3]. Fortunately, diagonalization provides a convenient way to take these powers[4].

So, in order to deal with (3), we would like to diagonalize the matrix

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix}.$$

[3] A successive-squaring approach would reduce the number of matrix multiplications needed to $O(\log n)$, though a closed-form solution is better than that.

[4] Recall tha a diagonalization of a (square) matrix A is a pair of matrices D and P such that D is diagonal, P is invertible, and

$$A = PDP^{-1}.$$

The spectral theorem tells us that such a factorization exists if and only if the linear transformation corresponding to A has a basis of eigenvectors. If it does, then we can let P be the change-of-basis matrix from that basis, on which the transformation acts as a diagonal matrix. So we get the desired decomposition, where D is a diagonal matrix with its entries the eigenvalues, and P has as its columns the eigenvectors.

To do this, we compute the eigenvectors of A by finding the characteristic polynomial:

$$\begin{aligned} c_A(\lambda) &= \det(A - \lambda I) \\ &= \det \begin{bmatrix} 2-\lambda & 3 \\ 1 & -\lambda \end{bmatrix} \\ &= (2-\lambda)\cdot(-\lambda) - 1\cdot 3 \\ &= \lambda^2 - 2\lambda - 3 \\ &= (\lambda-3)(\lambda+1) \end{aligned} \tag{4}$$

which has roots 3 and -1. Note that our characteristic polynomial computed as (4) is the same as the one we referred to as the characteristic polyomial of our recurrence in (1).

Now, we find eigenvectors corresponding to these eigenvalues. Recall that for any eigenvalue λ we are looking for a basis for $\operatorname{Ker}(A - \lambda I)$. So we compute

$$A - 3I = \begin{bmatrix} 2-3 & 3 \\ 1 & 0-3 \end{bmatrix} = \begin{bmatrix} -1 & 3 \\ 1 & -3 \end{bmatrix}$$

There are of course infinitely many nonzero vectors in the kernel of this matrix. We just need one, as the kernel is one-dimensional, so take $[3, 1]$. Similarly,

$$A - (-1)I = \begin{bmatrix} 2-(-1) & 3 \\ 1 & 0-(-1) \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 1 & 1 \end{bmatrix}$$

Again, we could take any nonzero vector in the kernel of this matrix; we will use $[-1, 1]$. That gives us a diagonalization

$$\begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix}^{-1}.$$

This will allow us to compute a closed form for a_n. By using this diagonal-

This makes taking powers of A easy, as

$$\begin{aligned} A^k &= (PDP^{-1})^k \\ &= \underbrace{PDP^{-1}PDP^{-1}\dots PDP^{-1}}_{k \text{ copies}} \\ &= P\underbrace{D\dots D}_{k \text{ copies}}P^{-1} \\ &= PD^kP^{-1} \end{aligned}$$

So we've translated the problem of taking powers of A to one of taking powers of D. Taking powers of diagonal matrices is simple, as it corresponds to taking powers of the diagonal entries. So we have our closed-form solution for A^k.

ization in (3) we get

$$\begin{aligned}
\begin{bmatrix} a_n \\ a_{n-1} \end{bmatrix} &= \begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix}^{n-1} \cdot \begin{bmatrix} 8 \\ 0 \end{bmatrix} \\
&= \left(\begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \right)^{n-1} \cdot \begin{bmatrix} 8 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}^{n-1} \cdot \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 8 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3^{n-1} & 0 \\ 0 & (-1)^{n-1} \end{bmatrix} \cdot \frac{1}{4} \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 8 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 3^n & (-1)^n \\ -3^{n-1} & (-1)^{n-1} \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -2 \end{bmatrix} \\
&= \begin{bmatrix} 2 \cdot 3^n - 2 \cdot (-1)^n \\ 2 \cdot 3^{n-1} - 2 \cdot (-1)^{n-1} \end{bmatrix}
\end{aligned}$$

so we've deduced the general form for our nth term, namely

$$a_n = 2 \cdot 3^n - 2 \cdot (-1)^n.$$

We can see that this is identical to the solution obtained by the method in the foreword. It is instructive to compare this with the first-order linear recurrence relation $a_n = r \cdot a_{n-1}$. In this case, we are multiplying by k each time, so we get a factor of r^n. Here we are 'multiplying' by a more complicated expression, namely the matrix A, but the solution is analogous.

3 Generating functions

This section is meant to give a gentle introduction to the idea of generating functions and their power in solving counting problems. These ideas are not limited to the solutions of linear recurrence relations; the provided references contain a little more information about the power of these techniques.

Our linear recurrence relation has a unique solution, which is a sequence of integers $\{a_0, a_1, a_2, \dots\}$. Given this information, we can define the *(ordinary) generating function* $A(x)$ of this sequence:

$$A(x) = \sum_{i=0}^{\infty} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots$$

We pause to consider this construction. $A(x)$ is a function of x, but our variable x is a dummy variable. We are concerned more so with this function's power series than the values of the function itself; in fact, we do not worry over-much about when this power series even has a positive radius of convergence[5]. But, since we have a sequence of coefficients, we can talk about a power series with those coefficients.

Now, we would like to use the recurrence condition $a_n = 2a_{n-1} + 3a_{n-2}$ to say something about the function $A(x)$. Now, one of the nice facts about our generating function is that multiplication by x will shift all the coefficients over by 1:

$$\begin{array}{lcccccccccccc} A(x) & = & a_0 & + & a_1 x & + & a_2 x^2 & + & a_3 x^3 & + & a_4 x^4 & + & \cdots \\ x \cdot A(x) & = & & & a_0 x & + & a_1 x^2 & + & a_2 x^3 & + & a_3 x^4 & + & \cdots \\ x^2 \cdot A(x) & = & & & & & a_0 x^2 & + & a_1 x^3 & + & a_2 x^4 & + & \cdots \end{array}$$

Next, we notice that the recurrence relation tells us that we can subtract twice the second row and three times the third row from the first row. Then all the coefficients will cancel, except for the first couple:

$$\begin{aligned} A(x) - 2xA(x) - 3x^2 A(x) &= a_0 + (a_1 - a_0)x \\ &\qquad + (a_2 - 2a_1 - 3a_0)x^2 \\ &\qquad + (a_3 - 2a_2 - 3a_1)x^3 \\ &\qquad + (a_4 - 2a_3 - 3a_2)x^4 \\ &\qquad + \cdots \\ &= a_0 + (a_1 - a_0)x + 0x^2 + 0x^3 + 0x^4 + \cdots \\ &= a_0 + (a_1 - a_0)x \end{aligned} \tag{5}$$

By our initial conditions, we know that $a_0 = 0$ and $a_1 = 8$. So we can plug those into (5) to get

$$A(x) \cdot (1 - 2x - 3x^2) = 8x$$

which means

$$A(x) = \frac{8x}{1 - 2x - 3x^2} \tag{6}$$

So, we've created a function whose power series at 0 is exactly $a_0 + a_1 x + \cdots$. Now, we just have to find out what that power series is explicitly. Fortunately, we know how to reduce complicated rational functions into simpler ones: we can use partial fractions. To do so, we need to factor the denominator of (6).

We can do this like so:

$$\begin{aligned} 1 - 2x - 3x^2 &= 1 + x - 3x - 3x^2 \\ &= (1 + x) - 3x(1 + x) \\ &= (1 + x)(1 - 3x) \end{aligned}$$

[5] Though in fact it does have a positive radius of convergence. Since the sequence $\{a_n\}$ is increasing, we have

$$a_n = 2a_{n-1} + 3a_{n-2} \leq 5a_{n-1}.$$

So, by the ratio test, the radius of convergence is at least 1/5. In fact, the radius is 1/3: see the exercises.

Or, since we just need to find the roots (and 0 is not a root by inspection), we can make the substitution $y = 1/x$ to get

$$\begin{aligned} 0 &= 1 - 2/y - 3/y^2 \\ &= y^2 - 2y - 3 \\ &= (y-3)(y+1) \end{aligned}$$

so the roots are $1/3$ and -1. Note the similarity between this polynomial and the characteristic polynomial (1).

So we can write (6) as

$$A(x) = \frac{8x}{1 - 2x - 3x^2} = \frac{c_1}{1+x} + \frac{c_2}{1-3x}$$

Cross-multiplying gives

$$8x = c_1(1-3x) + c_2(1+x)$$

which we can plug in $x = 1/3$ and $x = -1$ to find

$$\begin{aligned} 8/3 &= c_2(1 + 1/3) \\ -8 &= c_1(1+3) \end{aligned}$$

so $c_2 = 2$ and $c_1 = -2$. We now have a partial fraction decomposition of our generating function $A(x)$:

$$A(x) = \frac{-2}{1+x} + \frac{2}{1-3x}$$

Now the only thing left to do is to recognize both of the terms of this decomposition as instances of the geometric series formula[6]. That means

$$A(x) = (-2)\sum_{i=0}^{\infty}(-x)^i + (2)\sum_{i=0}^{\infty}(3x)^i$$

$$\sum_{i=0}^{\infty}(2 \cdot 3^i - 2 \cdot (-1)^i)x^i$$

By definition, a_n is just the nth coefficient in this power series, and now we know what that coefficient is:

$$a_n = 2 \cdot 3^n - 2 \cdot (-1)^n$$

So we've recovered a formula for the nth term. Notably, we only used $A(x)$ in the abstract; though it was useful to find what $A(x)$ was, the actual function is somewhat ephemeral. But our techniques allow us to translate recurrence problems into problems about operations on power series, and those are often easier to work with,

[6]Recall that

$$\frac{1}{1-t} = 1 + t + t^2 + \cdots$$

This can be shown by taking the limit of the partial sums, which are given by the finite geometric sum formula (proved by induction or factoring).

4 Analogies to ODEs

Recurrence relations are, in a certain sense, the discrete analogues of ordinary differential equations. Suppose we have a degree-2 linear constant-coefficient differential equation such as

$$f''(x) + af'(x) + bf(x) = 0$$

In applied mathematics, it is often the case that finding an exact answer to a complicated differential equation is practically impossible. In that case, one computational technique (the *method of finite differences*) is to pick a sequence of points $\{v_i\} \in \mathbb{R}$ and find values of the function that at least approximate the differential equation at those points. Let's approximate $f(v_i)$ with w_i. If the granularity is sufficiently high, then one can hope that the resulting approximation is close to the actual value of $f(v_i)$.

Now the problem is that, if we are considering only the values of the function at the $\{v_i\}$, we can no longer compute the exact values of the first and second derivative at those points. So we need to do some estimation. Let's assume the $\{v_i\}$ are evenly separated with a distance of h. Then we can make the following approxmations:

$$\begin{aligned} f'(v_i) &\approx \frac{f(v_{i+1}) - f(v_{i-1})}{2} \approx \frac{w_{i+1} - w_{i-1}}{2h} \\ f''(v_i) &\approx \frac{f(v_{i+1}) - 2f(v_i) + f(v_{i-1})}{h^2} \approx \frac{w_{i-1} - 2w_i + w_{i+1}}{h^2} \end{aligned}$$

(Both of these approximations come from using Taylor's theorem, and we can show that the error is bounded by a constant times the third derivative of f). If we plug this into our differential equation, we get that our approximations should satisfy

$$\frac{w_{i-1} - 2w_i + w_{i+1}}{h^2} + a\frac{w_{i+1} - w_{i-1}}{2h} + bw_i = 0$$

which we can rearrange to get

$$w_{i+1} = \frac{(2/h^2 - b)w_i + (\frac{a}{2h} - \frac{1}{h^2})w_{i-1}}{\frac{a}{2h} + \frac{1}{h^2}}$$

a linear recurrence relation! So we can approximate solutions to ODEs by using these reurrence relation ideas. Of course, in the constant-coefficient case we didn't need the help, but this idea is true for more complicated equations.

So it should come as no surprise that both of the methods described above are similar in nature to ideas from ODE theory. Recall that for a second-degree ODE, one possible trick is to define a new function $g = f'$ to turn the ODE into a system of first-degree ODEs. For instance, the ODE above would become

$$\begin{aligned} g'(x) + ag(x) + bf(x) &= 0 \\ f'(x) &= g(x) \end{aligned}$$

where we've now isolated g' and f'. We can write this as a linear system now:

$$\begin{bmatrix} g' \\ f' \end{bmatrix} = \begin{bmatrix} -a & -b \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} g' \\ f' \end{bmatrix}$$

This looks very similar to the matrix from the diagonalization example in Section 2, and the way to solve this system would be to diagonalize the matrix.

Similarly, we often solve (or "solve") a differential equation by constructing a Taylor series solution. We can postulate that our solution has a Taylor series solution that looks like

$$\sum_{n=0}^{\infty} c_n x^n$$

and plug this into our ODE to get some expressions for the coefficients. This is similar to the generating function approach.

5 Exercises

Here are a range of exercises having to do with the previously discussed material. The first few are direct applications of the methods - you can try solving them with either method - and then several problems extend the methods to fill out some details.

1. $a_n = 7a_{n-1} - 10a_{n-2}$, $a_0 = 4$, $a_1 = 11$
2. $a_n = -6a_{n-1} - 8a_{n-2}$, $a_0 = 2$, $a_1 = -2$
3. $a_n = 5a_{n-1} + 6a_{n-2}$, $a_0 = 1$, $a_1 = 6$
4. $a_n = 6a_{n-1} - 3a_{n-2} - 10a_{n-3}$, $a_0 = 3$, $a_1 = 6$, $a_2 = 30$
5. $a_n = 10a_{n-1} - 35a_{n-2} + 50a_{n-3} - 24a_{n-4}$, $a_0 = 0$, $a_1 = 1$, $a_2 = 2$, $a_3 = 3$
6. Consider the recurrence relation $a_n = 4a_{n-1} - 4a_{n-2}$. (Why is this different from the previous ones?)
   (a) Try both of the methods on this relation. What happens? (In the diagonalization case, you may have to argue about the powers of some other matrix).
   (b) What would happen if you had a triple-repeated solution?
   (c) Show that if there are no repeated roots, the matrix will always be diagonalizable.
   (d) Solve the recurrence relation $a_n = -3a_{n-1} + 4a_{n-3}$, $a_0 = 2$, $a_1 = 4$, $a_2 = 3$
7. Prove that the characteristic polynomial of the relation always shows up when using either method. In particular, show that if the roots are all unique, then the matrix is diagonalizable.
8. Show that in the generating function case, the radius of convergence of the power series is
   $$\frac{1}{|\lambda|}$$
   where λ is the root of the characteristic polynomial that is largest in absolute value.
9. (Extension). Recall that the Catalan numbers c_n are defined by the recurrence
   $$c_n = \sum_{i=0}^{n-1} c_i \cdot c_{n-i-1}.$$
   In this question, we derive the generating function for the Catalan numbers.

(a) Suppose we have a polynomial $p(x)$. We can write it as $\sum_{i=0}^{n} a_i x^i$. Show that

$$p^2(x) = \sum_{i=0}^{2n} \left(\sum_{j=0}^{i} a_j a_{i-j} \right) x^i.$$

(b) Let $C(x) = c_0 + c_1 x + c_2 x^2 + \ldots$ be the generating function of the Catalan numbers. Then use part (a) to prove

$$C^2(x) = \frac{1}{x}(C(x) - 1)$$

(c) Use the quadratic formula to recover $C(x)$. Note that there are two "roots"; compare

$$\lim_{x \to 0} C(x)$$

to pick the correct one.

(d) Find the radius of convergence (aka the distance to the closest problem point of $C(x)$), and conclude that c_n grows like 4^n.

6 References

- R. Johnsonbaugh, *Discrete Mathematics*.

  The course textbook for Math 61.

- M. Bona, *Introduction to Enumerative Combinatorics*.

  The course textbook for Math 184 has material on generating functions, including other types beyond the ordinary, and their uses.

- R. Stanley, *Enumerative Combinatorics* (volume 1).

  The classic work on enumerative combinatorics. It is available online at `http://www-math.mit.edu/~rstan/ec/ec1.pdf`, and is an extensive look at counting problems of all kinds.