Difference equation lecture notes

Paul Johnson <pauljohn@ukans.edu>.

July 15, 2011

The “ordinary” writings about difference equations are concerned about systems with constants for input, usually. Or the inputs are simple, step functions or impulse functions. We need to review some of that, then concentrate on the claims/findings of Hamilton, Time Series Analysis, Chapters 1 & 2.

1 Constant input

1.1 First order difference equation

Given the “ordinary” setup, we have:

$y_t = A * y_{t-1} + B$

B is constant input, A is a constant “multiplier”. This can be solved easily, just begin at y_0 and go over and over!

$y_1 = A * y_0 + B$

$y_2 = A(A * y_0 + B) + B = A^2 y_0 + AB + B$

$y_3 = A(A^2 y_0 + AB + B) + B = A^3 y_0 + A^2 B + AB + B$

Could this possibly get more dull. I'm sick of it already. Easily I can see the pattern, can't you?

$$y_t = A^t y_0 + A^{t-1} B + A^{t-2} B + A^{t-3} B + \ldots AB + B \quad (1)$$

$$y_t = A^t y_0 + B[A^{t-1} + A^{t-2} + A^{t-3} + \ldots A + 1] \quad (2)$$

Recall the geometric series we discussed in class. In the brackets, that's what we have, and the series can be solved explicitly as

$A^{t-1} + A^{t-2} + A^{t-3} + \ldots A + 1 = \frac{1-A^t}{1-A}$

So put that into equation 2 and you get

$$y_t = A^t y_0 + B \frac{1-A^t}{1-A}$$

or

$$y_t = A^t y_0 + B \frac{1}{1-A} - B \frac{A^t}{1-A} = A^t y_0 + \frac{B}{1-A} \left[1 - A^t\right]$$

or

$$y_t = A^t[y_0 - \frac{B}{1-A}] + \frac{B}{1-A} \tag{3}$$

This is what is known as a **solution** of the original difference equation. It is a solution because it gives a formula for y_t that does not depend on y_{t-1} . You can look at this in any of these ways, which ever gives you insight.

I often forget to emphasize that this solution is true only if $A \neq 1$. But now I'm emphasizing it. If $A = 1$, then it is impossible to divide anything by $(1-A)$. Instead, the solution ends up as:

$$y_t = A^t y_0 + B[A^{t-1} + A^{t-2} + A^{t-3} + \ldots A + 1] = y_0 + (t-1) * B \tag{4}$$

This is a symptom of the time-series research problem of the "unit root." There is a dramatic change in the behavior of the dynamic system if $A = 1$.

1.2 Properties

It is evident in 3 that the path of y_t depends on the value of A. Assuming $A \neq 1$, then:

If $A \geq 1$, then y_t "explodes", getting bigger and bigger.

If $0 \leq A < 1$, then y_t gradually "moves toward" $\frac{B}{1-A}$

The value $\frac{B}{1-A}$ is a vital value, it is the "equilibrium" level of the system.

If $-1 < A < 0$, then y_t , it goes toward $\frac{B}{1-A}$ but it does not do it peacefully. Instead, it "oscillates". Oscillates means it alternates, going above and below the equilibrium.

If $A < -1$, then y_t "explodes", and it oscillates too. $|y_t|$ gets bigger and bigger.

Draw some pictures of this!

A dynamical system is said to be stable if, as $t \to \infty$, then y_t remains bounded. It does not "explode". Either it tends to a constant value, as in the AR(1) case, or it oscillates within a limited region.

1.3 Look at the "big picture"

Suppose I said to you that the equation

$$y_t = Ay_{t-1} + B$$

has a solution if $A \neq 1$, a general solution, like this:

$$y_t = k_1 * \lambda^t + k_2$$

That is true, which means that when you are confronted with the difference equation, then you should think to youself: "aha! this a linear time system and there is some λ which determines the dynamics of the system.“ In the terms of the trade, the term λ is known as a **root.**

The only problem is to find out what λ might be. Now, if you look at the previous section, you easily know that

$$\lambda = A$$

and

$$k_1 = [y_0 - \frac{B}{1-A}]$$

$$k_2 = \frac{B}{1-A}$$

It does not seem so complicated to me.

If you read a big fat book on differential or difference equations, you will see talk about a general solution and a particular solution. The general part refers to solving the homogeneous equation:

$$y_t - Ay_{t-1} = 0$$

and then adding a particular solution for the other part of the original model, the "input" B.

1.4 The giant picture

Now, suppose you said your system had more lagged inputs, as in

$$y_t = A * y_{t-1} + B * y_{t-2} + C \tag{5}$$

From looking at this, I can tell you that, for "most" values of A and B there is a general solution with 2 "roots", λ_1 and λ_2 of the form

$$y_t = k_1\lambda_1^t + k_2\lambda_2^t \tag{6}$$

Now, you want to know the values of λ_1 and λ_2 if you want to know if this system is stable. Its not so easy to get those "roots" as it is in the previous case, but at least we understand our goal. We want to know those roots.

Goldberg contends (p. 134) that these roots are found as the solution of the "auxiliary equation", also sometimes known as the **characteristic equation**. To derive that, set C=0, so we consider only the heart of the system:

$$y_t - Ay_{t-1} - By_{t-2} = 0$$

Because we get tired of fussing over minus signs, relabel the coefficients as $a_1 = -A$ and $a_2 = -B$.

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} = 0$$

Goldberg says it is "easy" to see there is at least one "test solution" like

$$y_t = \lambda^t$$

Insert that "test solution" into 5 and you get:

$$\lambda^t + a_1 * \lambda^{t-1} + a_2 * \lambda^{t-2} = 0$$

Which you can rearrange (by dividing everything by λ^{t-2}) as

$$\lambda^2 + a_1\lambda + a_2 = 0 \tag{7}$$

This is the **characteristic equation** for the second order dynamic system**.** The possible solutions are referred to as λ_1 and λ_2 . If these two solutions are not the same, then the general solution I claimed in 6 is valid. That is the case of "distinct roots".

Now, we want to know if the roots are:

real valued: smaller then 1 in absolute value, we want to know if they are positive or negative

or

complex valued: the system oscillates, perhaps without bound!

Systems with complex roots make my head hurt, because then I have to remember what $\sqrt{-1}$ means. And sin() and cos() and modulus and other stuff I hate.

Worry about that another later.

1.5 The impossibly huge picture

Just remember this:

In general, if your discrete time system has M lags, then its solution has M roots, and your problem is to find out about their values.

It is important to find because:

1) You can check system "stability"

2) You can see dynamic properties (oscillation)

3) You can interpret the importance of inputs into the system. Certain values of the roots will cause the system to respond strongly to new input. Other cases might not.

The search for roots is what drives us to work on the so-called characteristic equation. It is characteristic of the system's underlying dynamics. Given a theory like this:

$$y_t = A * y_{t-1} + B * y_{t-2} + C$$

write the part with the y's like so:

$$y_t - A * y_{t-1} - B * y_{t-2} = 0$$

And we know there will be a general solution which fits into a pattern like 6. Of course, if you have a really big difference equation, like:

$$y_t = A * y_{t-1} + B * y_{t-2} + C * y_{t-3} + D * y_{t-4} + E$$

then the homogeneous general part of this is

$$y_t - A*y_{t-1} - B*y_{t-2} - C*y_{t-3} - D*y_{t-4} = 0$$

And if we know the general part of the solution is a set of constants k_j and λ_j such that

$$y_t = k_1\lambda_1^t + k_2\lambda_2^t + k_3\lambda_3^t + k_4\lambda_4^t$$

Now, how in the world do we get the coefficients? We want those roots! We must have them.

It turns out that there is a big fat theorem which says that the roots are found as the solution to the so-called "characteristic equation". I've been confused sometimes because I have seen this written down two different ways. Sometimes it goes like this

$$\lambda^4 - A\lambda^3 - B\lambda^2 - C\lambda - D = 0 \tag{8}$$

Other times people get rid of the minus signs (the way I did in the second order case) or they convert the whole equation by dividing everything through by a new variable $z = \frac{1}{\lambda}$.

If you can solve that for the λ 's, then you have the roots, and you can analyze them to determine the system's behavior. There are different ways to solve for these magical numbers, and we will get to that later in these notes, kinda.

Also note one more thing. A polynomial like the characteristic equation can also be factored, and it turns out the roots pop up in an obvious way:

$$\lambda^4 - A\lambda^3 - B\lambda^2 - C\lambda - D = (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3)(\lambda - \lambda_4) \tag{9}$$

In Mickens's book *Difference Equations*, on p. 124, he states as a theorem that, as long as the roots are distinct, then the solution to the homogeneous difference equation with p lags takes the general form:

$$y_t = c_1\lambda_1^t + c_2\lambda_2^t + c_3\lambda_3^t + \ldots + c_p\lambda_p^t$$

1.6 Now, about that "unit circle" business.

In the first order equation, we know the "stability condition" is that the coefficient $|\lambda| < 1$. In the higher order equations, a similar equation applies for all the individual λ_j . If all the λ_j are real numbers, then the stability requirement is just that each one is smaller than one, or $|\lambda_j| < 1$.

Unfortunately, sometimes, the root can be a complex number. For example, in a second order system, the characteristic equation is a quadratic equation, and since 9th grade or so we know the famous formula to solve quadratic equations. The "roots" of

$$ax^2 + bx + c = 0$$

are given by the formula

$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

When I was in school, we said that a solution only existed if $b^2 - 4ac > 0$. Back then, they taught us "you can't take the square root of a negative number."

Well, we aren't in high school anymore! This is college. You can do it! We must solve all these characteristic equations, no matter what the coefficients. And that means figuring out what it means if there is a square root of a negative number! That's what complex numbers are for. In complex number theory (not my specialty, for sure), there is a thing i defined as the square root of minus 1:

$$i = \sqrt{-1}$$

and it turns out that we can take the square root of $\sqrt{b^2 - 4ac} > 0$ if we use that i thing.

A complex number has two parts, a real part and an imaginary part. Generally speaking, a complex number looks like so:

$$n_1 + n_2 * i$$

n_1 is the "real part" and the rest is the "complex part," a number n_2 is a real number multiplied by the square root of -1.

In the above example of the quadratic equation, suppose that b=2, a=1 and c =3. Then the root is

$$x = \frac{2 \pm \sqrt{4-12}}{2} = \frac{2 + \sqrt{-8}}{2} = 1 \pm \frac{\sqrt{-1} * \sqrt{4}\sqrt{2}}{2} = 1 \pm i\frac{2\sqrt{2}}{2}$$

$$x = 1 \pm i\sqrt{2}$$

The real part is 1, the imaginary part is $\sqrt{2}$.

So, if you have a characteristic equation with complex roots, you have some trouble ahead of you, but not all that much. To understand all the details, it involves sin(), cos(), and the pythagorean theorem. You can study that on you own if you want, but the only really vital thing is this.

Stability a complex root requires the following: $\sqrt{n_1^2 + n_2^2}$

What's the idea behind this? If λ_j is complex, then raising it to powers like λ^2 and λ^3 and so forth may either cause the value to shrink to zero or it might instead explode to infinity. The condition for the value to shrink to zero is that the value of the modulus, of "length", of the complex number must be smaller than one. Look at Hamilton, p. 709. The modulus is defined by the pythagorean theorem, i.e.,

$R = \sqrt{n_1^2 + n_2^2}$

I don't have the patience to explain the steps that it takes to justify this claim.

If you took a Cartesian plan, a basic graph, you can find all the values of n_1 and n_2 for which the modulus is smaller than one. If you dres that, then you would have a circle centered at zero, with a radius of one. This **unit circle** is the thing people are referring to when they say that a system is stable if all the roots are inside the unit circle.

With real valued roots, inside the unit circle just means $|\lambda_j| < 1$, but with complex roots, it is a little more complicated to understand.

2 What if the inputs are not constant?

Here's where I've always gotten stuck in the past! The Hamilton Chapters 1 & 2 cleared up a lot of questions for me.

2.1 First-order difference equation

Think about the B in the first order difference equation for a minute. It is a constant input. But imagine that, instead of a constant B, at each time step we get a different number. Follow Hamilton and use the variable w_t to refer to the different numbers for inputs.

Who cares? Then the first order difference equation described above in 1 on page 1 only needs a little bit of adaptation. I like the letter A, but nobody else does, they all seem to like the Greek ϕ , so I cave in and start using that. Let's follow Hamilton and suppose a first order system:

$$y_t = \phi y_{t-1} + w_t. \tag{10}$$

Hamilton starts iterating this equation at t=0, supposing he knows the value at time -1 for y, which he calls y_{-1} . So we have this:

$$y_0 = \phi y_{-1} + w_0$$

plug that in go get y_1 , y_2 , and so forth, until we have:

$$y_t = \phi^{t+1} y_{-1} + \phi^t w_0 + \phi^{t-1} w_{t-1} + \phi^{t-2} w_{t-2} + \ldots + \phi w_{t-1} + w_t \tag{11}$$

The key thing is that we are looking at the "inputs" w as just a string of numbers. Instead of adding B at every timestep, we are just adding on w_t . The numbers $\{w_0, w_1, w_2, \ldots, w_t\}$ are just numbers, nothing special. Just numbers that get added.

2.1.1 Dynamic Multiplier

Well, maybe they are not just numbers. They are "inputs." They are quantities we theorize about, things that we think might affect y. Hamilton is often interested in the "**dynamic multiplier**", the impact on y_t caused by a change in one of the input values at one time.

In the above example 11, the first order model, it is painfully obvious. The impact of a change in w_t depends on "how long ago" it was and also on these coefficients ϕ . If you could somehow "grab" the variable w_0 and make it one unit larger, then it is apparent that the change in y_t would be ϕ^t . If you have to be skeptical about it, look at it like this. Add one unit to w_0 and rewrite 11:

$$y_t^{new} = \phi^{t+1} y_{-1} + \phi^t (w_0 + 1) + \phi^{t-1} w_{t-1} + \phi^{t-2} w_{t-2} + \ldots + \phi w_{t-1} + w_t \tag{12}$$

Now if you subtract 11 from 12, you get:

$$y_t^{new} - y_t = \phi^t (w_0 + 1) - \phi^t w_0$$

$$y_t^{new} - y_t = \phi^t \tag{13}$$

This shows the discrete change, the gap between the original y_t and y_t^{new} .

To the "high powered" math types, its not very interesting to look just at the discrete change, they want to look at derivatives or partial derivatives. Maybe we need to talk in more detail about that.

Hamilton uses partial derivatives to describe this effect.

$$\frac{\partial y_{t+j}}{\partial w_t} = \phi^j \tag{14}$$

In other words, a change in w at time t has an effect j time periods later that is equal to ϕ^j .

Please note that the interpretation of the coefficient ϕ in this finding is strikingly similar to the first section above (see equation 3). If the system is "stable", meaning $|\phi| < 1$, then the effect of a change in w has a smaller and smaller effect as time goes by. If the system is unstable, then the effect of w gets bigger and bigger.

So, unlike the boring, old interpretation of difference equations, rather than just caring about stability, we are not concerned about the "long lasting impact" of a variable input into the system.

2.1.2 Effect of a permanent increase in w_t .

The claim in 14 shows the effect of a unit change in w_t at any one time. We can use that insight to wonder about the impact of a "permanent" change in w . If, beginning at some time t, the value of w_t is increased, and all following $w's$ are increased similarly, then the effect at t+j is the sum of the effects of all the $w's$:

$$1 + \phi + \phi^2 + \phi^3 + \phi^4 + \phi^5 + \ldots \phi^j$$

If you let j get really big, say, infinite, and $|\phi| < 1$, then this impact in j periods after the change is:

$$\frac{1 - \phi^{j+1}}{1 - \phi}$$

and when j is huge and ϕ^j tends toward 0, so as a result the impact of a permanent change in w_t is

$$\frac{1}{1 - \phi} \tag{15}$$

Do you see why its a rather silly question to measure the impact of a change when ϕ is 1 or greater? See why it is so vital that a dynamic system be "stable" if we are going to get anything useful out of it?

2.2 Higher order difference equations

Add more lagged y's, and a coefficient ϕ_j for each one:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_{p-1} y_{t-p-1} + \phi_p y_{t-p} + w_t. \tag{16}$$

If you follow along with Hamilton (p. 8-9), you start to see his plan. Use whatever mathematical tool you can to find out the dynamic properties of the system, and, in particular, measure the impact of changes in w_t .

On p. 8, Hamilton introduces a way to convert a single equation model into a matrix equation. That's not done just because matrices are cool, but also because they have powers and there are lots of results about them. His attention ends up focusing on the matrix F, which is:

$$F = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \vdots & 0 & 0 \\ 0 & 1 & 0 & \vdots & 0 & 0 \\ 0 & 0 & 1 & \vdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & 1 & 0 \end{bmatrix} \tag{17}$$

You arrive at that matrix by writing the main equation you are interested in in the first row, and then add a set of identities like

$$y_k = y_k \tag{18}$$

where k={y-1,y-2,...,y-p}.

Look at the top of p. 9 in Hamilton. You see that if we repeat the iteration process, putting in new inputs and calculating y_t at each step, it is as though we are multiplying F by itself, repeatedly.

We can talk details, but the big news is tha the impact of a change in w_t ends up being filtered through the matrix F and F^2 and F^3 and so forth.

3 What's that Eigenvalue thing all about?

Oh, please, not that. I took linear algebra in 1985. You expect me to remember that? I didn't like it then, much. If I have to do it again, at least I want to know what for!

3.1 Justification of awful pain and suffering

There are 2 reasons why the eigenvalue comes back at me after all these years.

1. Eigenvalues of the matrix F end up solving the characteristic equation 8. In fact, the eigenvalues *are* the roots of the characteristic equation. (Hamilton proves this in Proposition 1.l, p. 10).

2. Eigenvalues of the matrix F end up giving ingredients in the calculation of $\frac{\partial y_{t+j}}{\partial w_t}$. The dynamic multiplier is a linear sum of the eigenvalues, something like $c_1\lambda_1 + c_2\lambda_2$ for a second order system, and the coefficients c_1 and c_2 have formula that include the roots as well.

3.2 Actual pain and suffering

Suppose you have some square matrix. I don't know why we can't call it F. The definition of an eigenvalue goes like this. A number λ is an eigenvalue if the determinant $|F-\lambda I|=0$. The determinant of a 2x2 matrix is so easy that calculating it is like falling out of bed. If the matrix is:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Then

$$\left|\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right| = ad - cb$$

A determinant of a bigger matrix is more difficult to calculate, usually involves some high order polynomials.

So if the matrix that you are trying to get a determinant is $F-\lambda I$, that means you are getting the determinant of

$$F-\lambda I = \begin{bmatrix} \phi_1-\lambda & \phi_2 & \phi_3 & \dots & \phi_{p-1} & \phi_p \\ 1 & -\lambda & 0 & \vdots & 0 & 0 \\ 0 & 1 & -\lambda & \vdots & 0 & 0 \\ 0 & 0 & 1 & \vdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \vdots & 1 & -\lambda \end{bmatrix} \tag{19}$$

Solving the determinant $|F-\lambda I|$ usually means solving a polynomial in λ .

There can be up to as many different solutions–values of λ –as there are rows in this matrix. And finding may not be easy. In fact, I don't think there is a general formula for solving these if the order is more than 3.

Now here are 2 mathematical complications:

1. If all of the solutions are distinct, some of the mathematics works out more easily.

2. If all of the solutions are real-valued, then we are more familiar with interpreting the results.

Note Hamilton has a lot of effort expended to the case in which the eigenvalues are distinct, versus the case in which some are repeated. Note he also has a pretty big effort devoted to working out the details of the complex solution cases.

3.3 Characteristic. Schmarasteristic.

Now, as far as solving the characteristic equation goes, this just works like magic. You can adjust the coefficient λ to make $|F-\lambda I|=0$, and all such numbers are the roots of the characteristic equation. Now days, we even have computer smart enough to find the λ_j for us :)

3.4 Remember we wanted the dynamic multiplier?

As far as finding the dynamic multipliers, it is a little more work.

Hamilton, on p. 11, uses a result from linear algebra for matrices with distinct eigenvalues. The result is that

$$F = T\Lambda T^{-1}$$

for some matrix T and where Λ is a matrix that collects up the eigenvalues, like so:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \tag{20}$$

Now there is a stunning thing about this result. If you need to calculate Λ^2 is is just the squared items from Λ . If you need Λ^3 , it is just the cubed elements of Λ . Its just as simple as pie!

Furthermore, if you calculate F^2 , it ends up being as simple as $T\Lambda^2 T$, and if you need F^3 it is just $T\Lambda^3 T^{-1}$.

So, it is trivially easy to tell the values of F^t .

That means, whenever we need a number from F^t, it is no trouble to get it.

The multiplier is, in general, a combination of the eigenvalues. The impact of a change in w_t that will be felt after j periods is like so:

$$\frac{\partial y_{t+j}}{\partial w_t} = c_1\lambda_1^j + c_2\lambda_2^j + ... + c_p\lambda_p^j \tag{21}$$

There's a formula on p. 12 in Hamilton to use for the c coefficients.

So we know that the dynamic multiplier, which describes the impact of a unit change in w_t is a weighted combination of the eigenvalues.

3.5 Didn't you want the impact of a permanent change too?

But wait! it gets even better than that. We can use the eigenvalues as a "safety check" to make one more very important observation. Look back up at equation 15. We want a result like that, but for equations with more lagged y's.

The impact of a permanent change in w_t in a higher order equation can be seen as a direct extension of 15. That is, assuming that ALL roots (eigenvalues) are inside the unit circle, it is meaningful to make the exact same simplifications that led to 15 in a bigger model, and as a result, the impact of a permanent one unit change in w_t after many many (actually, an infinite number of) periods is:

$$\frac{1}{1 - \phi_1 - \phi_2 - \phi_3 - \cdots - \phi_p}$$

(See Hamilton's p. 20 result, Proposition 1.3). Please don't forget this is only valid if the system is stable.

4 What the heck is that L thing doing in my reading again?

L is the lag operator, and it has some nice properties. Unfortunately, not many authors clearly explore the operator. Instead, they tend to just want to throw it around as if it were a number, which works much of the time, but not always. One of the strength's of Hamilton's chapter 2 is a pretty thorough explanation of what L is and what is good for.

4.1 What are we sure is true of L?

I don't know about a comprehhensive list, but:

1. definition:

$$x_{t-1} = Lx_t$$

2. raise L to powers

$$x_{t-2} = L(Lx_t) = L^2 x_t$$

$$x_{t-3} = L(L^2 x_t) = L^3 x_t$$

3. L obeys linearity, so you can multiply by a constant:

$$4 * Lx_t = L(4x_t)$$

4. L obeys a distributive law. You can also act as if L is a coefficient and do things like:

$$L(3 * x_t + 4 * y_t) = 3Lx_t + 4Ly_t$$

or

$$L(1 + 3L) = L + 3L^2$$

5. polynomial grouping is meaningful. You can do "factoring" and "multiplying" of expressions involving L

$$6L^2 + 5L + 1 = (3L + 1)(2L + 1)$$

Then we know it is the same thing to apply either the left or the right to a variable x_t . So, first apply the left hand side:

$$(6L^2 + 5L + 1)x_t = 6x_{t-2} + 5x_{t-1} + x_t$$

and that is the same as applying the right hand side:

$$(3L + 1)(2L + 1)x_t = (3L + 1)(2x_{t-1} + x_t)$$

$$= 6Lx_{t-1} + 3Lx_t + 2x_{t-1} + x_t$$

$$= 6x_{t-2} + 5x_{t-1} + x_t$$

What do we think is not true of L? Remember L is an operator, so you can't assume it always works like a real number. So, for example, it meaningful to talk about the inverse of

$$(1 - \phi L)$$

only if $|\phi| < 1$. As Hamilton observes, in that case the inverse $(1 - \phi L)^{-1}$ exists and the idea of "dividing" something by $(1 - \phi L)$ make sense. (Hamilton, p. 28; yes, its just another example of the geometric series).

4.2 Think of a difference equation as a polynomial in L

Take any discrete time system, like 11:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_{t-p} y_{t-p} + w_t \tag{22}$$

If you use the L operator, this is

$$y_t = \phi_1 L y_t + \phi_2 L^2 y_t + ... + \phi_{t-p} L^p y_t + w_t \tag{23}$$

And you might as well write:

$$(1 - \phi_1 L - \phi_2 L^2 - ...\phi_p L^p) y_t = w_t \tag{24}$$

If we want a "solution" for y_t then we want y_t on the left hand side, all by itself, an we wish we could do something simple like:

$$y_t = \frac{w_t}{(1 - \phi_1 L - \phi_2 L^2 - ...\phi_p L^p)} \tag{25}$$

The big problem is that we can't write such a thing down because we can't just bash L about as if it were a number. In some particular cases, Hamilton shows (p. 30) that it is meaningful. In particular, guess what: it depends on the eigenvalues. Again, eigenvalues, I can't stand it.

Still, the notion that we just write use shorthand like

$$C(L) = 1 - \phi_1 L - \phi_2 L^2 - ...\phi_p L^p \tag{26}$$

and

$$y_t = C^{-1}(L) w_t$$

is appealing. Note how doing this makes it clear that the the current value of y_t is a weighted combination of inputs! That's exactly what we wanted for the "dynamic multipliers" model. If only we knew if $C^{-1}(L)$ were a meaningful thing, and how to calculate it!

4.3 The result is especially clear in a first order difference equation.

Hamilton shows a simple case of an AR(1) model, one for which $C(L) = 1 - \phi L$, and he shows that, by simple algebra, that we can get what we want. Start with

$$(1-\phi L)y_t = w_t$$

and multiply both sides by (1+$\phi L + \phi^2 L^2 + \phi^3 L^3 + ... + \phi^t L^t$). You are allowed to do that with L's, as we described above.

Then evaluate this by doing the multiplication term by term:

$$(1+\phi L + \phi^2 L^2 + ... + \phi^t L^t)(1-\phi L)$$

What do you end up with?

$$(1-\phi^{t+1}L^{t+1}) \tag{27}$$

Man, oh man. that means:

$$(1-\phi^{t+1}L^{t+1})y_t = (1+\phi L + \phi^2 L^2 + ... + \phi^t L^t)w_t \tag{28}$$

which means

$$y_t - \phi^{t+1}L^{t+1}y_t = (1+\phi L + \phi^2 L^2 + ... + \phi^t L^t)w_t \tag{29}$$

And since $L^{t+1}y_t$ is just the value of y at t = -1, then the second term on the left hand side is:

$$\phi^{t+1}L^{t+1}y_t = \phi^{t+1}y_{-1}$$

Then move that to the right hand side of the big equation, and look what we have:

$$y_t = \phi^{t+1}y_{-1} + (1+\phi L + \phi^2 L^2 + ... + \phi^t L^t)w_t \tag{30}$$

This got a lot of work done! We have y_t by itself on the left hand side, and some "other stuff" on the right, just what we wanted! Now, we see y_t by itself on the left hand side, it means we found the "practical equivalent" of $C^{-1}(L)$ =$(1-\phi L)^{-1}$. It does "almost" what we want, except there is little problem of the term

$$\phi^{t+1}L^{t+1}y_t = \phi^{t+1}y_{-1}$$

That is "extra", "unwanted","hated", "undesirable", and generally ugly. **But, if we assume that ϕ <1, then we can assert that this extra part "shrinks" to zero, and we throw it away.**

The conclusion is that, if $\phi < 1$, then we can act *as if* $(1-\phi L)^{-1}$ exists, and that means we can write things like

$$(1-\phi L)y_t = w_t$$

$$(1-\phi L)^{-1}(1-\phi L)y_t = (1-\phi L)^{-1}w_t$$

$$y_t = (1-\phi L)^{-1}w_t$$

$$y_t = (1+\phi L + \phi^2 L^2 + ... + \phi^t L^t)w_t \tag{31}$$

Please remember, it is only "as if" we are allowed to divide both sides by $(1-\phi L)$. Since L is not a number, it is not strictly meaningful to speak of division by L .

4.4 What if that "sneaky trick" worked more generally?

Consider a pth order difference equation:

$$y_t - \phi_1 L y_t - \phi_2 L^2 y_t - \ldots - \phi_p L^p y_t = w_t \tag{32}$$

or

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) y_t = w_t \tag{33}$$

When I was in 9th grade, I don't think I was paying attention, but since then I've learned it is true that if you can write a polynomial by factoring it. That means there are some numbers $\lambda_1, \lambda_2, \ldots, \lambda_p$ such that:

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) = (1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L)$$

So that means we can substitute that for the left hand side of 33, and we have:

$$(1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L) y_t = w_t \tag{34}$$

Now, we saw in the previous section that we **do** know of a way to invert things like this. We could employ the trick from the previous section to get $(1 - \lambda_1 L)^{-1}$. We apply that rule over and over again, and our problems are solved! We end up with y_t on the left hand side, all by itself. All we need are the coefficients $\lambda_1, \lambda_2, \ldots, \lambda_p$

4.5 Then this new kind of characteristic equation happens

Recall the characteristic equation 8. We said the system is stable if the roots are all inside the unit circle.

Now, when people write the difference equation with L's in it, they arrive at a different kind of equation that looks almost just like a characteristic equation. Look at 23 and notice there is an equivalent of the characteristic equation, except a little different.

If you replace the lag operator L with the real number z, then the polynomial in L looks like

$$1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0 \tag{35}$$

This is just the same as the old characteristic equation, except now we have replaced λ by $\frac{1}{z}$. If we talk about the roots of this equation in z, we are talking about the same roots that we had in the other equation.

But the stabilty conditions are reversed. So, if the original characteristic equation required that all roots be inside the unit circle, what does this new equation say about the roots of z? They have to be outside the unit circle!

Note the very excellent paragraph Hamilton, (p. 32), where he mentions the frequent confusion when some authors talk about roots inside or outside the circle without precisely describing what equation they are talking about. Wow. That really answered some questions I had accumulated.

4.6 Now, back to the dynamic multipliers again.

Look at this equation in the factored polynomial above. Supposing you did apply the inverse for each term, you would end up with y_t on the left hand side, all by itself.

$$y_t = (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}\cdots(1-\lambda_p L)^{-1} w_t \tag{36}$$

Doing this requires that each of the roots $\lambda_1, \lambda_2, ...\lambda_p$ is inside the unit circle.

Now, if you recall what it "really" means to apply $(1-\lambda_1 L)^{-1}$, what we are really doing is multiplying by a long sum, (1 + $(1+\lambda_1 L+\lambda_1^2 L^2+...+\lambda_1^t L^t)$. We have to do that for $(1-\lambda_2)^{-1}$ and so forth. At the end, on the right hand side we have all kinds of L 's and λ 's floating around. We don't care to actually write all that out, we might as well note, however, that the formula would have to be something like:

$$y_t = \psi_t + \psi_1 w_{t-1} + \psi_2 w_{t-2} + ... + \psi_t w_0 \tag{37}$$

The coefficients ψ_t might be algebraically complicated, but we know for sure they depend on the λ_i .

These coefficients ψ_t are just the dynamic multipliers! Hamilton p. 35 gives the formulae, I'm too tired for that now.