Mathematics for Economists
Aalto BIZ
Spring 2021
Juuso Välimäki

Difference equations

Difference equations are perhaps the simplest example of dynamical systems. In a dynamical system the state of the system at point n in time, x_t, and a system equation can be used to extrapolate the entire future for the system, i.e. the values of the state x_{t+k} for all $k \in \mathbb{N}$. We will discuss the cases where $\boldsymbol{x}_t \in \mathbb{R}$ as well as the case where $x_t \in \mathbb{R}^n$. Except for some motivating examples, we shall deal with linear systems where the system equation takes the form

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t$$

for some $n \times n$ matrix $\boldsymbol{A}$.

Motivating examples

Fibonacci sequence

Let's start with some examples. The most famous difference equation is the one giving rise to the Fibonacci sequence. We set $x_0 = 0$, $x_1 = 1$ and for all $t > 2$,

$$x_t = x_{t-1} + x_{t-2}.$$

You can compute by repeatedly substituting previous values the sequence $0, 1, 1, 2, 3, 5, 8, 13, ...$ In some sense, the system equation is already a solution to the problem. Nevertheless it is not easy to see what x_{200} is or even to know its approximate size. By a solution to a difference equation, we mean here a formula that depends on the initial values x_0 and t so that $x_t = g(x_0, t)$ for some function g. In fact, we can look at the problem as one where the function g is the unknown to be solved. We will see how to compute the values for x_t in the Fibonacci sequence.

Compartmental models of infectious diseases

The most widely used models for the spread of infectious diseases in a population is the SIR-model, where the population at each point in time t is partitioned into susceptible (S_t), infected (I_t) and recovered (R_t). The susceptible become infected if they are in contact with infected with some probability β. The infected recover (or die) at rate γ.

This gives rise to a three-variable system:

$$\begin{aligned} S_{t+1} &= S_t - \beta S_t I_t, \\ I_{t+1} &= I_t + \beta S_t I_t - \gamma I_t, \\ R_{t+1} &= R_t + \gamma I_t. \end{aligned}$$

This system does not have an explicit solution for the variables as a function of time. It is almost linear, but the number of infections depends on the meetings between susceptible and infected and therefore involves the product of these two variables. Nevertheless, the system is simple enough to simulate for various values of β, γ and many of the scenarios offered by THL for the COVID -pandemic have in fact just been extrapolations of this model for different parameter values.

The most important determinant of the spread of the disease is the epidemiological threshold $r = \frac{\beta S_t}{\gamma}$. This measures the average number of new infections spread by an infected agent before recovery. If this number falls below 1, then the spread of the disease stops. At the start, the disease spreads (assuming that $\beta > \gamma$), but as a larger fraction of the susceptible population becomes infected and recovered, herd immunity is reached and the spread stops.

Solow growth model

The Solow growth model is perhaps the most important motivating example from economics. In this model, labor is kept fixed at L over time and capital K_t changes over time as a result of savings. The aggregate production function is $y_t = F(K_t, L_t)$ and it is usually assumed to be an increasing concave, linearly homogenous (constant returns to scale) so that

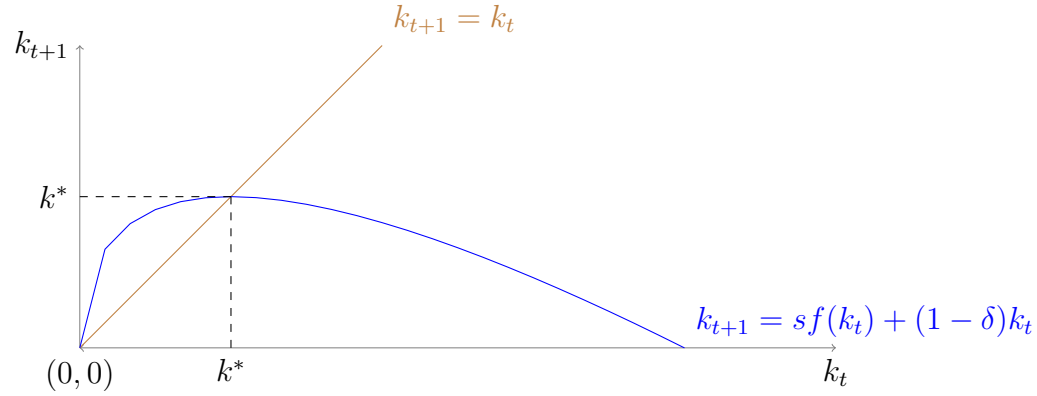
$$F(K_t, L_t) = L_t F\left(\frac{K_t}{L_t}, 1\right) := L_t f(k_t),$$

where $k_t := \frac{K_t}{L_t}$ is the pe capita capital stock. It is often assumed that $\lim_{k\to 0} = \infty$ and $\lim_{k\to\infty} = 0$. Let's assume here that $y_t = k_t^\alpha$ for some $0 < \alpha < 1$.

The output y_t is divided between savings and consumption. The assumption is that a constant fraction sy_t is saved. As you can recall from Principles of Economics II, savings equals investment and investment goes into next period's capital. Capital depreciates at rate δ per period. Taking all this together, we get

$$k_{t+1} = sk_t^\alpha + (1-\delta)k_t.$$

Let's draw the graph of $sk_t^\alpha - \delta k_t$ in the (k_t, k_{t+1}) plane with the 45-degree line. Of special interest is the point k^* such that $s(k^*)^\alpha - \delta k^* = k^*$. If you start the system with $k_0 = k^*$, the system stays there forever since $k_1 = sk_0^\alpha - \delta k_0 = k_0$ and therefore also $k_t + n = k_0$ for all n. We call k^* the steady state or a rest point of the system.



We can picture the movement of the system by positing an initial point k_0 on the horizontal axis. If $k_0 < k^*$, then $k_1 = f(k_0) > k_0$. You can locate the k_1 on the horizontal axis by reflecting on the 45-degree line. Repeating this process, you can show that from any initial point, k_t converges to k^* as $t \to \infty$. We say that k^* is a globally stable steady state.

You may want to note that

$$k^* = (\frac{s}{1+\delta})^{\frac{1}{1-\alpha}}.$$

The steady state capital level is determined by the savings rate s for a fixed technology (fixed α and δ). Hence countries that save more grow more.

You will see more sophisticated versions of the model in your intermediate macroeconomics course.

Markov process

A population consists of three income classes $i \in \{1, 2, 3\}$. If you are in class i, your children are in income class j with probability p_{ji}. Let P be the matrix with a typical element p_{ij}.

Let $\boldsymbol{x}_0 = \mathbf{e_i}$ if you are in class i. Then the probability that you child is in class j is given by the column vector

$$\boldsymbol{x}_1 = P\boldsymbol{x}_0 = P\mathbf{e_i}.$$

But then the probability that your grandchild is in class j is given by the column vector

$$\boldsymbol{x}_2 = P\boldsymbol{x}_1 = P^2\boldsymbol{x}_0,$$

and in general,

$$\boldsymbol{x}_{t+1} = P\boldsymbol{x}_t.$$

This is a linear difference equation system with constant coefficients and we will see how to solve such systems. The interesting questions here include: How does the probability distribution of your sixth successor generation depend on your income class? In other words, how much social mobility is there in the society.

Linear difference equations with constant coefficients

The simplest form of difference equations are linear difference equations with constant coefficients. These can be written as:

$$\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + \boldsymbol{b}_t,$$

where b_t is a given sequence. If $b_t = 0$ for all t, we have a homogenous equation. We start with the simplest homogenous equations where $x_n \in \mathbb{R}$ and $A = a \in \mathbb{R}$.

Solving the homogenous equation is very easy. If $x_{t+1} = ax_t$ for all t, then $x_{t+k} = a^k x_t$. Hence any sequence of the form $x_t = ca^t$ solves the

difference equation. If we are given the initial value x_0, the solution is $x_t = x_0 a^t$. In other words, the initial value pins down the coefficient c of the general solution.

Consider next an inhomogenous equation,

$$x_{t+1} = ax_t + b,$$

where $b_t = b$ for all t. Clearly the constant solution $x_t = \frac{b}{1-a}$ for all t solves the equation. I claim that also $x_t = ca^t + \frac{b}{1-a}$ solves the equation. But this follows immediately from the fact that $ca^{t+1} = aca^t$.

This principle holds more generally. If you have a particular solution x_t^P to the inhomogenous equation and the general solution of the homogenous equation x_t^H, then the general solution to the problem is $x_t^P + x_t^H$. This is called the principle of superposition and it arises from the linearity of the equations in x_{t+1}, x_t. It is valid also for the case with $\boldsymbol{x}_t \in \mathbb{R}^n$.

Consider next linear systems with constant coefficients. Let $\boldsymbol{x}_t \in \mathbb{R}^n$ for all t and let $\boldsymbol{A}$ be an $n \times n$ matrix of real numbers. A linear homogenous system is then given by:

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t.$$

As before, we can 'solve' this by repeated substitution to get

$$\boldsymbol{x}_{t+k} = \boldsymbol{A}^k \boldsymbol{x}_t.$$

Hence I could write the general solution as $\boldsymbol{x}_t = \boldsymbol{A}^t c$ for some vector $c = (c_1, ..., c_k)$. I do not consider this a real solution since it is almost impossible to see what $\boldsymbol{A}^t$ is except in some very special cases.

If $\boldsymbol{A}$ is a diagonal matrix with diagonal elements $a_1, ..., a_n$, then the solution becomes

$$x_{i,t} = ca_i^t \text{ for } i \in \{1, ..., n\}.$$

Here we have essentially independent variables and the difference equation for each can be solved separately.

Eigenvalues and eigenvectors

To deal with the general case, we want to change the basis in $\mathbb{R}^n$ so that $\boldsymbol{A}$ is diagonal in that basis. This involves the eigenvectors and eigenval-

ues of $\boldsymbol{A}$. You can visualize the effect of matrix multiplication on vectors as consisting of two operations: a rotation and a stretching or shrinking. Eigenvectors of $\boldsymbol{A}$ are those vectors that are not rotated, i.e. if $\boldsymbol{x} \neq 0$ is an eigenvector of $\boldsymbol{A}$, then for some $\lambda \in \mathbb{R}$,

$$\boldsymbol{A}x = \lambda \boldsymbol{x}.$$

We may write this more compactly as

$$(\boldsymbol{A} - \lambda I)\boldsymbol{x} = 0,$$

where I is the $n \times n$ identity matrix. But from basic linear algebra, we know that a homogenous linear equation can have a non-zero solution only if the matrix does not have full rank, i.e. if $\det(\boldsymbol{A} - \lambda I) = 0$. The values of λ for which this determinant is zero are called the eigenvalues of $\boldsymbol{A}$.

The determinant of $(\boldsymbol{A} - \lambda I)$ is called the characteristic polynomial of $\boldsymbol{A}$ so the eigenvalues are the roots of the characteristic polynomial. If $\boldsymbol{A}$ has n distinct eigenvalues $\lambda_1, ..., \lambda_n$, then it has also n linearly independent eigenvectors $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ so that

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i.$$

Let' see an example on how to compute the eigenvalues and vectors. Let

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then

$$\boldsymbol{A} - \lambda I = \begin{pmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{pmatrix},$$

and

$$\det(\boldsymbol{A} - \lambda I) = \lambda^2 - \lambda - 1.$$

We have $\det(\boldsymbol{A} - \lambda I) = 0$ if

$$\lambda_1 = \frac{1+\sqrt{5}}{2}, \lambda_2 = \frac{1-\sqrt{5}}{2}.$$

The corresponding eigenvectors are:

$$\boldsymbol{v}_1 = (\frac{1+\sqrt{5}}{2}, 1), \boldsymbol{v}_2 = (\frac{1-\sqrt{5}}{2}, 1).$$

In the lecture slides, you will see a connection of this example and the Fibonacci sequence in the motivating examples.

A useful thing to keep in mind about eigenvalues is that the sum of the eigenvalues equals the trace of the matrix and the product of the eigenvalues equals the determinant of the matrix. This is particularly useful for inference about the signs of eigenvalues.

Since the characteristic polynomial may fail to have real roots, eigenvalues correspond to the case where the matrix does not have any directions that are not rotated. To see an easy example of such a matrix, consider the 90-degree rotation anticlockwise:

$$\boldsymbol{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The characteristic polynomial for this matrix is $\lambda^2 + 1$ which obviously does not have real roots. If the eigenvalues are complex numbers, the eigenvectors are also have complex coordinates. We do not have time in this course to pursue this, but it should be pointed out that the method outlined below for solving the difference equations extends also to the case with complex eigenvalues.

It should not come as a surprise that rotations rotate all vectors. You may want to pursue the geometric implications for a 3x3 matrix with a single real eigenvalue and hence a single real eigenvector.

Eigenvectors, eigenvalues and difference equations

I can express any $\boldsymbol{x} \in \mathbb{R}^n$ given in the usual coordinate system in the coordinate system of spanned by the eigenvectors by simple matrix multiplication. Let $\boldsymbol{V} = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \dots \boldsymbol{v}_n]$ be the matrix formed by the eigenvectors. Then for any vector $\boldsymbol{y}$ expressed in the coordinate system of the eigenvectors, we can translate it to the standard system by $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{y}$. Similarly any x in the standard system is $\boldsymbol{y} = \boldsymbol{V}^{-1}\boldsymbol{x}$ in the system of the eigenvectors.

$$\boldsymbol{y}_{t+1} = \boldsymbol{V}^{-1}\boldsymbol{x}_{t+1} = \boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{x}_t = \boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V}\boldsymbol{y}_t.$$

Now we want to show that $\boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V} = \Lambda$, where Λ is the diagonal matrix of eigenvalues. But this is the same claim as (premultiply by $\boldsymbol{V}$):

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\Lambda.$$

But this follows immediately from the fact that $\boldsymbol{V}$ consists of the eigenvectors of $\boldsymbol{A}$.

Hence we have: $\boldsymbol{y}_t = (y_{1,t}, ..., y_{k,t}) = (c_1\lambda_1^t, ..., c_n\lambda_n^t)$. Since $\boldsymbol{x}_t = \boldsymbol{V}\boldsymbol{y}_t$, we have the general solution:

$$\boldsymbol{x}_t = c_1\lambda_1^t\boldsymbol{v}_1 + ... + c_n\lambda_n^t\boldsymbol{v}_n.$$

Note that $\boldsymbol{A}^k = \boldsymbol{V}\Lambda\boldsymbol{V}^{-1}$. Therefore we could have also concluded that

$$\boldsymbol{x}_t = \boldsymbol{V}\lambda^k\boldsymbol{V}^{-1}\boldsymbol{x}_0.$$

The two methods give the same results since $\boldsymbol{V}\boldsymbol{c} = \boldsymbol{x}_0$ or $\boldsymbol{c} = \boldsymbol{V}^{-1}\boldsymbol{x}_0$.

Sometimes a matrix has a repeated eigenvalue. Consider for example

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Then the characteristic equation is $(1-\lambda)^2 = 0$ and the matrix has a single eigenvalue $\lambda = 1$ and therefore a single eigenvector $(1, 0)$. This matrix cannot be diagonalized in the procedure that we had above. Luckily enough all matrices $\boldsymbol{A}$ can be expressed as:

$$\boldsymbol{A} = Q^{-1}BQ,$$

where Q is a matrix of generalized eigenvalues and B is upper triangular. Since the powers of upper triangular matrices are easy to compute, the same procedure as before can be applied for solving the model. See the book for the details on this.

Properties of the solutions

For all (homogenous) systems of linear difference equations, 0 is a steady state. If $\boldsymbol{A}$ has full rank, it is the only steady state. Does the system eventually converge to its steady state?

Look at the general solution

$$\boldsymbol{x}_t = c_1\lambda_1^t\boldsymbol{v}_1 + ... + c_n\lambda_n^t\boldsymbol{v}_n.$$

If $|\lambda_i| < 1$ for all i, then $\boldsymbol{x}_t \to 0$ for all c. We say that in this case, the origin is a globally stable steady state or a sink. If $|\lambda_i| > 1$ for all i, then

the length of $\boldsymbol{x}_t$ grows without bound for all $c \neq 0$. We say that the origin is unstable or a source. Finally if If $|\lambda_i| < 1$ for some i and If $|\lambda_i| > 1$ for some i, then the length of $\boldsymbol{x}_t$ grows without bound if $c_i \neq 0$ for some i with $|\lambda_i| > 1$. If $c \neq 0$ only for i with $|\lambda_i| < 1$, then $\boldsymbol{x}_t$ converges to the origin. In this last case, we say that origin is a saddle point for the system. If $\lambda_i = 1$ for some i, then origin is neither stable, unstable nor a saddle.

Example: Linear endogenous growth model

Consider the following linear system in two variables. I will explain the economic content of the model below.

$$\begin{aligned} c_{t+1} &= \beta\rho c_t, \\ k_{t+1} &= \frac{1}{1+n}(\rho k_t - c_t), \end{aligned}$$

or

$$\begin{pmatrix} c_{t+1} \\ k_{t+1} \end{pmatrix} = \begin{pmatrix} \beta\rho & 0 \\ \frac{-1}{1+n} & \frac{\rho}{1+n} \end{pmatrix} \begin{pmatrix} c_t \\ k_t \end{pmatrix},$$

where $\rho = 1 - \delta + A$.

Since the system equation is given by a lower triangular matrix $\boldsymbol{A}$, we see immediately that the eigenvalues of $\boldsymbol{A}$ are the diagonal elements $\beta\rho$ and $\frac{\rho}{1+n}$. An eigenvector corresponding to eigenvalue $\beta\rho$ is:

$$(\rho(1 - (1+n)\beta), 1),$$

and an eigenvector for eigenvalue $\frac{\rho}{1+n}$ is $(0, 1)$.

Hence, we have:

$$\boldsymbol{A} = \begin{pmatrix} \beta\rho & 0 \\ \frac{-1}{1+n} & \frac{\rho}{1+n} \end{pmatrix} = \boldsymbol{V}^{-1}\Lambda\boldsymbol{V},$$

where

$$\Lambda = \begin{pmatrix} \beta\rho & 0 \\ 0 & \frac{\rho}{1+n} \end{pmatrix}, \boldsymbol{V} = \begin{pmatrix} \rho(1 - (1+n)\beta) & 0 \\ 1 & 1 \end{pmatrix}.$$

Exercise: Show that:

$$\boldsymbol{V}^{-1} = \begin{pmatrix} \frac{1}{\rho(1-(1+n)\beta)} & 0 \\ -\frac{1}{\rho(1-(1+n)\beta)} & 1 \end{pmatrix}.$$

We have shown that for initial conditions (c_0, k_0), the system is solved by:

$$\begin{pmatrix} c_t \\ k_t \end{pmatrix} = \begin{pmatrix} (\beta\rho)^t c_0 \\ \frac{(\beta\rho)^t c_0}{\rho(1-(1+n)\beta)} + (\frac{\rho}{1+n})^t \left(k_0 - \frac{c_0}{\rho(1-(1+n)\beta)}\right) \end{pmatrix}.$$

Optional: Economics of the model

Consumers in an economy decide in each period how much to consume and how much to save in the form of capital. Additional consumption of Δ (small) units in period t brings marginal utility $u'(c_t)\Delta$. By saving until $t+1$, the consumer will get $\beta(1-\delta+r_t)u'(c_{t+1})\Delta$, where β is the discount factor and r_t is the interest rate and $(1-\delta)$ is the amount of capital remaining in $t+1$ after depreciation δ. At an optimal consumption path, the consumer cannot gain by reallocating marginal consumptions across periods. Hence we have the *Euler equation*:

$$u'(c_t) = \beta(1-\delta+r_t)u'(c_{t+1}).$$

In a competitive economy, the interest rate is given by the marginal product of capital. If $f(k_t)$ is the production function in the economy measured in terms of capital per worker k_t, then:

$$r_t = f'(k_t).$$

Finally, capital accumulates according to (similar to the Solow model above):

$$k_{t+1} = \frac{1}{1+n}\left(f(k_t) + (1-\delta)k_t - c_t\right),$$

where n is the growth rate of the working population. If we assume that $u(c_t) = \ln(c_t)$ and $f(k_t) = Ak_t$, then we have the linear model that we saw above.

In more advanced courses in macroeconomics, you would learn that a reasonable condition on the model parameters is that $\beta(1+n) < 1$. In this case, you can see by dividing k_t by c_t that the only choice of c_0 such that consumption c_t remains positive and does not vanish relative to k_t is that:

$$c_0 = k_0(\rho(1-(1+n)\beta)).$$

Capital stock is normally viewed as a predetermined variable and that argument above fixes the initial consumption at such a level that consumption remains proportional to the capital stock throughout. Depending on the parameter values, this system may display sustained economic growth (if $\beta\rho > 1$).

Linearizing non-linear systems

For your future information, I note here that if x^* is a steady state of a nonlinear system, we can use Taylor's first order approximation to analyze the local behavior of the system around the steady state (you'll do this in macroeconomics a lot). Here is the idea.

Suppose that $\boldsymbol{x}_{t+1} = f(\boldsymbol{x}_t)$ and $\boldsymbol{x}^* = f(\boldsymbol{x}^*)$. Then we have

$$\boldsymbol{x}_{t+1} = f(\boldsymbol{x}_{t+1}) = f(\boldsymbol{x}^*) + D_{\boldsymbol{x}} f(\boldsymbol{x}^*)(\boldsymbol{x}_t - \boldsymbol{x}^*) \text{ or}$$

$$\boldsymbol{x}_{t+1} - \boldsymbol{x}^* = D_{\boldsymbol{x}} f(\boldsymbol{x}^*)(\boldsymbol{x}_t - \boldsymbol{x}^*).$$

But this is a linear system in the deviations from the steady state and we can apply the analysis from the linear case in the for small deviations. You can classify the steady states of nonlinear models locally as we just did for the linear system (but globally). Just look at the absolute values of the eigenvalues and compare to 1.

Markov model

Consider the system

$$\boldsymbol{x}_{t+1} = P\boldsymbol{x}_t$$

for a stochastic matrix P, i.e. non-negative matrix whose elements in each column sum up to 1. You have already shown in Problem set 0 that $\lambda = 1$ is an eigenvalue for all Markov matrices.

It can be shown that in the case with strictly positive entries, all other eigenvalues are less that one in absolute value. Therefore x_t converges in the long run to the eigenvector (whose coordinates are normalized to sum to 1) corresponding to eigenvalue 1.

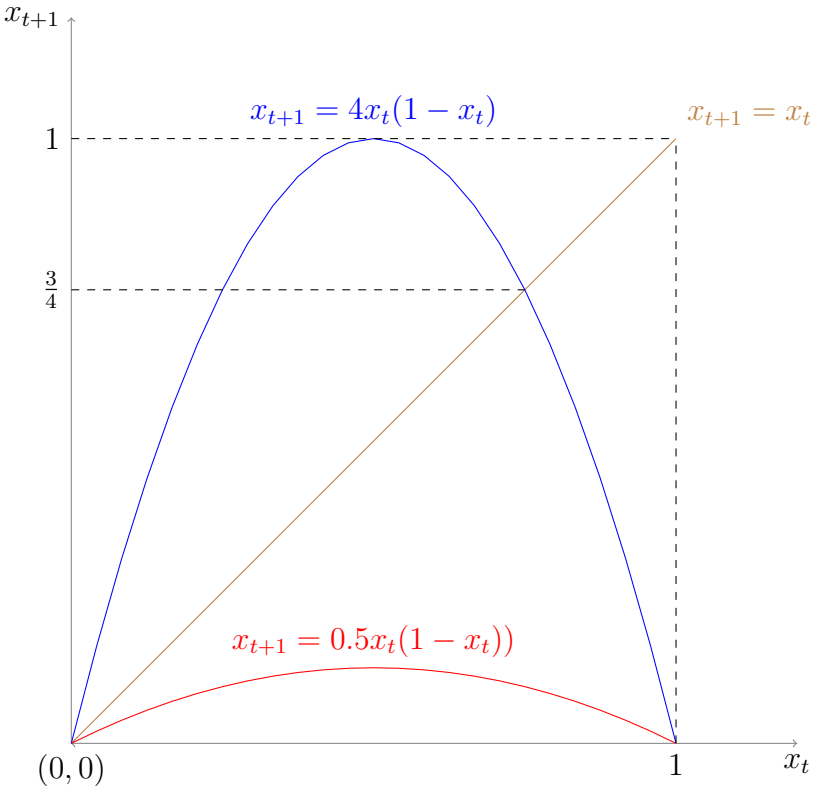
The second largest (in length) eigenvalue measures the speed of convergence to this eigenvector.

Extra: Non-linear systems

General difference equations for $x_t \in \mathbb{R}$ look deceptively simple, but can give rise to really surprising behavior. One famous example is the logistic equation on the unit interval: choose $0 < x_0 < 1$ and compute for $t > 1$:

$$x_{t+1} = r x_t (1 - x_t).$$

This is a nice differentiable function whose values remain in $(0, 1)$ for all t as long as $r < 4$. To analyze a difference equation on the real line, the first step is to look at the graph of the system equation.

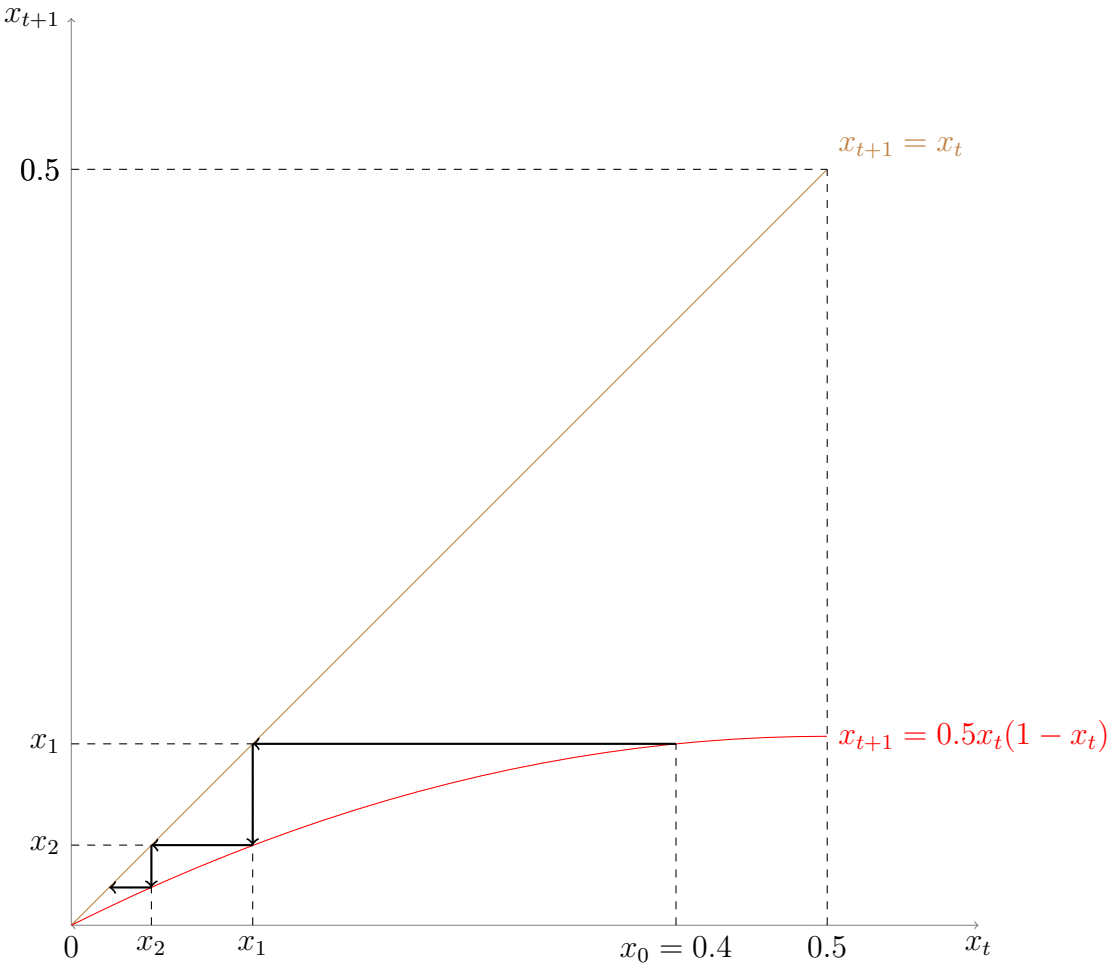


What is the significance of the intersections of $x_{t+1} = r x_t (1 - x_t)$ and the 45 -degree line $x_{t+1} = x_t$. The system stops at any such point, because if $x_{t+1} = x_t$, then also $x_{t+k} = x_t$ by repeated substitution into the system equation. These are called the steady states of the dynamical system.

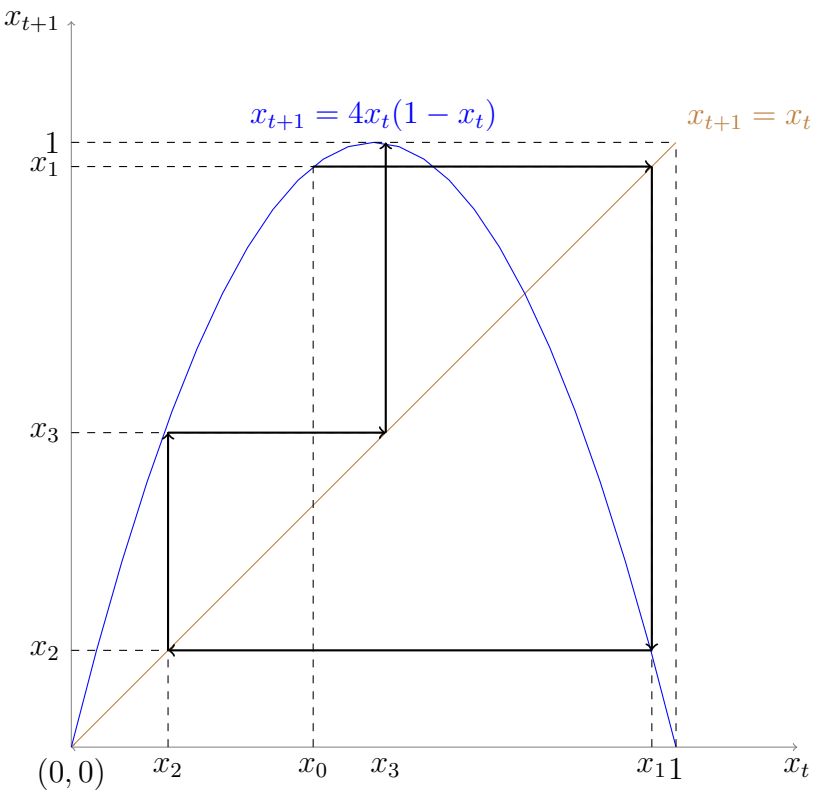
Notice that the system has a single steady state at $x = 0$ if $r < 1$ (can you show this?). For $4 > r > 1$, the system has another steady state at $x = \frac{r-1}{r}$. What happens to the values of x_t as t grows?

Here is a nice graphical way of seeing what happens to the sequence. Lets graph the function in a coordinate system where x_{t+1} is on the vertical and x_t is on the horizontal axis. Draw the graph of $x_{t+1} = f(x_t)$ and pick a starting point $x_0 = 0.4$ for example on the horizontal axis. You can read $x_1 = f(x_0)$ on the graph. You need to picture x_1 on the horizontal axis to see where x_2 is located. But you can do this by reflection through the 45 -degree line. The you just continue the procedure.

Lets look first at the case $x_{t+1} = \frac{1}{2}x_t(1 - x_t)$, i.e. lets take the red curve in the previous picture.



As you can see, for any starting x_0, the system x_t converges quite quickly to 0. If we take the blue graph from the first picture, things look quite different. Let's follow the system again for a few rounds starting at $x_0 = 0.4$.



It is much harder to see where the system might converge and actually the long run behavior of this simple dynamical system is very complicated. In fact, one can show that the system has cycles of all lengths, i.e. for all k, you can find $x_0, x_1, ..., x_{k-1}, x_k = x_0$ such that $x_j + 1 = 4x_j(1 - x_j)$ for all $j \in 0$.

This model is so simple that you can simulate it very easily with Excel. You can see for example how quickly the paths from nearby starting values diverge.