

## 16. Lyapunov Functions and Equilibria

*The beauty of mathematics only shows itself to more patient followers.*

—Maryam Mirzakhani

In this chapter, we learn Lyapunov functions, which provide conditions for a model to achieve equilibrium. Lyapunov functions are real-valued functions defined over the configuration system that are indexed by time. In each time step, a Lyapunov function assigns a value to the configuration. If the configuration changes—that is, if the model is not at an equilibrium—then the Lyapunov function’s value decreases by a fixed amount. A Lyapunov function also has a minimum value, which means that eventually its value must stop decreasing. When that happens the model reaches an equilibrium. We can use Lyapunov functions to show, for example, why the local majority model converges.

The key takeaway from this chapter will be that if we can construct a Lyapunov function for a model, the model must go to an equilibrium. We cannot get a periodic orbit, randomness, or complexity. Even more, we can also bound the time to convergence to that equilibrium as we show when we construct a Lyapunov function for the local majority model.

The chapter has six parts. We first define Lyapunov functions and then apply them in the Race to the Bottom Game. We then construct Lyapunov functions for the local majority model and a model of ordering activities. In the fourth part, we show why we can construct Lyapunov functions for some exchange markets, and why we cannot for others. We then see why the Game of Life lacks a Lyapunov function as well. We then discuss a deceptively vexing mathematics problem that always goes to an equilibrium for which no Lyapunov function has been found. The chapter concludes by

returning to the question of whether equilibria are desirable.

## Lyapunov Functions

A *discrete dynamical system* consists of a space of possible *configurations*—think of these of as multidimensional states of the world such as an initial collection of live and dead cells in the game of the life—along with a *transition rule* that maps the configuration at time  $t$  into a configuration at time  $t + 1$ . A *Lyapunov function* maps configurations into the real numbers and satisfies two assumptions. First, if the transition function is not at an equilibrium, the value of the Lyapunov function falls by a fixed amount (more on that in a moment). And second, the Lyapunov function has a minimum value. If both assumptions hold, then the dynamical system must attain an equilibrium.

## Lyapunov Theorem

Given a discrete time dynamical system consisting of the transition rule  $x_{t+1} = G(x_t)$ , the real-valued function  $F(x_t)$  is a Lyapunov function if  $F(x_t) \geq M$  for all  $x_t$  and if there exists an  $A > 0$  such that

$$F(x_{t+1}) \leq F(x_t) - A \text{ if } G(x_t) \neq x_t.$$

If  $F$  is a Lyapunov Function for  $G$ , then starting from any  $x_0$ , there exists a  $t^*$ , such that  $G(x_{t^*}) = x_{t^*}$ , and the system attains an equilibrium in finite time.

We first construct a Lyapunov function within the *Race to the Bottom Game*, which captures strategic environments in which players choose levels of support such that each player prefers to provide just less than the average level.

## The Race to the Bottom Game

Each of  $N$  players proposes a level of support in  $\{0, 1, \dots, 100\}$  in each period. The player closest to  $\frac{2}{3}$  of the average level of support wins a prize in that period.

The game can be used to explain reductions in state government spending for social programs such as support for the indigent. No governor or state legislature wants to appear heartless. Yet none wants to offer lavish programs that attract indigent populations from neighboring states. Each state wants to offer some funding, but less than the average. The same incentives exist for countries choosing environmental regulations or tax rates. Countries prefer to be less restrictive on environmental policies and have lower than average taxes in order to attract business.

Whether the Race to the Bottom Game attains an equilibrium depends on the behavioral rules of the players. If, for example, players choose random levels of support, then outcomes will be random. Random levels would not make sense given the game's payoff structure. Here we assume the following behavioral rule that aligns with experimental findings.<sup>1</sup> In the first period, we assume that every player chooses a random level of support less than 50. Thereafter, each player chooses a level at least 1 less than  $\frac{2}{3}$  of the previous period's average. If that number is less than zero, then each player chooses zero.

It is straightforward to show that the maximum level of support from any player satisfies the conditions for a Lyapunov function. The maximum level of support has a minimum at zero. And in each period the maximum level of support falls by at least 1 given that levels of support take integer values. Thus, at some point, everyone proposes zero support. The players have raced to the bottom. In this example, the model produces an undesirable result. To prevent a race to the bottom requires changing the game. To increase support for the indigent, a federation could shift to federal funding or impose a floor on spending.<sup>2</sup>

As an aside, suppose that we allow players to choose any real number in the interval between zero and 100 rather than integer values. If in each round, players chooses a level of support equal to  $\frac{2}{3}$  of the previous mean, the average level of support will decrease over time but it will never attain the equilibrium of zero. As in Xeno's paradox, the process would get closer and closer to zero but never reach it. Thus, we must assume a minimal decrease ( $A$ ) to ensure an equilibrium.

## Equilibrium in the Local Majority Model

We now return to the local majority model. We define our Lyapunov function to be the *total disagreement* in the population: the sum over all cells of the number of neighboring cells in the opposite state. To prove the model attains an equilibrium, we must show that if a cell changes its state, then total disagreement falls by at least a fixed amount.

The algebra is not too complicated. First, if a cell changes its state, it must have been in a minority relative to its neighbors. We know that at least five of its neighbors were in the opposite state and at most three were in the same state. Therefore, when the cell switches states, the number of cells that disagree with the cell decreases by at least 2 (see [figure 16.1](#)). To calculate the change in total disagreement, we must add in the changes to total disagreement contributed by neighboring cells. The five or more cells that now agree with the cell's state have lower disagreement (by 1 each) and the three or fewer cells that previously agreed now have higher disagreement (by 1 each). Therefore, total disagreement across all neighboring cells falls by at least 4.

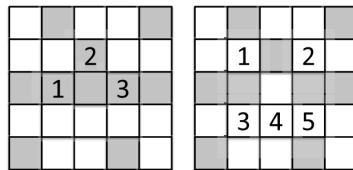


Figure 16.1: Total Disagreement Falls by 4 in Local Majority Model

We have therefore proven that even though some cells may have more disagreement, total disagreement satisfies the conditions for a Lyapunov function. The local majority model must, therefore, converge to an equilibrium—not just sometimes or most of the time, but all of the time. We also know the rate of convergence. Any time that a cell changes its state, total disagreement falls by at least 4. It follows that a configuration with total disagreement of 100 must reach an equilibrium within 25 periods. More generally, a configuration with total disagreement of  $D$  must reach equilibrium in  $\frac{D}{4}$  periods. As noted in [Chapter 15](#), the equilibrium reached will almost always be an inefficient pattern of splotches that includes

frustrated cells.



## Self-Organization: New York and Disney World

Our next application uses Lyapunov functions to prove the existence of an equilibrium in the *self-organizing activities model*. The model consists of a population of individuals and a set of activities that each individual can do during the day. The key assumption of the model will be that each person prefers less-crowded activities. Fewer people means less wait for the machines at the gym and a shorter line at the bakery or coffee shop. This model was motivated by a quote by Thomas Schelling from *Micromotives and Macrobehavior* in which he describes amazement at how cities self-organize—how traffic patterns, the flow of pedestrians, the number of people in parks and restaurants, and store inventories all reach appropriate levels with little central planning. How does the corner store always have four bottles of pure maple syrup from Cedarville, Michigan? How come the bakery runs out of fresh rye bread each day about twenty minutes before closing time? This order emerges even though the city's diverse actors—the tourists, store owners, residents, and delivery people—have limited information about the entire city.

## Self-Organizing Activities Model

A city offers  $A$  activities. Each day consists of  $L$  time periods. Each person in large population of size  $M$  chooses a **routine**, an order to participate in a set of  $L$  activities (out of a larger set of  $K$  possibilities) across the  $L$  time periods. A person's **congestion** equals the number of other people who choose the same activities as her at the same times.

To prove the model converges, we show that the *total congestion*, the sum of the congestion levels of the entire population, satisfies the conditions for a Lyapunov function. When a person lowers her congestion, she lowers her contribution to total congestion and also lowers the congestion by 1 for each person she no longer meets and raises congestion by 1 for each new person she encounters. Given that she lowers her own congestion, it follows that more people belong to the first group than the second. For example, suppose that a person was going to a crowded gym at 8:00 a.m. and a crowded coffee shop at 4:00 p.m. If she switches time slots and finds the coffee shop to be nearly empty at 8:00 a.m. and the gym to be less crowded at 4:00 p.m., she reduces congestion for herself and for all the people she met previously. She does increase congestion for the smaller number of people she now meets, but total congestion falls (and by at least 1). Given that total congestion cannot fall below zero, the system must reach an equilibrium.

Although, in general, we have no guarantee that the system will locate an efficient equilibrium, this model almost always converges to a configuration of almost minimal total congestion. In an inefficient configuration, more people choose one activity—say, going to the gym—during a period than some other activity—say, going to the coffee shop. If the difference in congestion is high at those two activities, an individual can lower her congestion by switching the times she goes to the gym and the coffee shop. If the coffee shop and the gym had equal numbers of people in that other period, the exchange would reduce total congestion.<sup>3</sup>

The model explains some of the order that we see in the world. It gives insight into how cities can self-organize to near-efficient configurations without central planners. It also tells us why amusement parks, such as Disney World, may not. Each day, Disney World has new attendees, who lack the time to try new routes. Without help from central planners, some attractions at Disney World would have huge lines, while others would have no wait. Disney World tries to limit these inefficiencies by allowing people to sign up for particular attractions at specific times and by having employees steer people to less crowded areas.

## Pure Exchange Economies

We can also use Lyapunov functions to explore when pure exchange economies attain equilibria and when they may not. A *pure exchange economy* consists of a set of consumers, each of whom has an endowment of goods as well as preferences. We might think of a set of people who show up at a marketplace or bazaar with something to trade with others, such as eggplants, cheese, or woven blankets. Each trade has a cost in time and effort for both parties. In order for two people to trade, each must benefit by an amount that exceeds this cost of trading.

Rather than construct a Lyapunov function that always decreases by a fixed amount and has a minimum, we do the opposite: we show that *total happiness* always increases by a fixed amount and has a maximum value. By assumption, any time two people trade, their happiness levels increase by at least the cost of trading. In addition, each person brought a fixed endowment of goods, so total happiness has a maximum value. The assumptions of a Lyapunov function are met, so the system goes to an equilibrium. At that equilibrium, the allocation need not be efficient. Of course, if it is not efficient, some people in the market might be able to identify a trade that makes them happier.

In constructing this argument, we assume that only people who participate in the trade derive happiness (or unhappiness). In other exchanges that might not be true. Imagine that Iraq promises to trade oil in return for nuclear weapons from Pakistan. The leaders of both countries may be happier, but total happiness, measured globally, would fall. The rest of the world might not be happy with Iraq building up a nuclear arsenal.

The impact felt by people in other countries is called a *negative externality*. When an exchange market includes negative externalities, trades need not raise total happiness. In our earlier example of a pure exchange market, when people trade fruits, vegetables, blankets, and tools, few externalities exist. The presence of externalities implies that we cannot say whether the system will reach an equilibrium. Trades in arms and oil such as the one described could beget other trades. Iraq's growing stockpile of nuclear weapons could lead Saudi Arabia to demand more military

support from its allies. This in turn could lead to actions by other countries in the region. Global happiness, or global security for that matter, may jump up and down with each action. We cannot be sure that trades will ever stop.

Whether or not Lyapunov functions exist in a trading context depends on the size of the negative externalities, as can be seen through an example told to me by a former undergraduate student. Her employer was moving to new offices that would include a large room with open desks where analysts would sit. Her manager proposed randomly assigning analysts to desks and then allowing people to trade where they sat. He thought this would lead to a good outcome because of a belief that free exchange produces efficiency.

My former student realized that even though any two people who traded desks would be happier, their former and new neighbors need not be. A person might feel hurt if a current neighbor, especially one he may have chosen to be near, traded to a desk across the room. The former neighbor might also not like the new neighbor, who might talk loudly on the phone. The former neighbor may then himself move. Relocations might continue for a long time. And each one might chip away at morale. The plan looked risky. The organization needed its employees to trust one another and treat one another with respect. Those behaviors would be hard to maintain in employees who kept having people trade to move away from them. When her manager thought through the model, he abandoned his plan.<sup>4</sup>

The story does not end there. This same manager had also purchased office chairs in a variety of styles and colors and had planned to randomly assign chairs and allow people to trade. In this instance, my student (using model thinking) told her manager to allow trading of chairs. Chair trades would not produce any externalities and would be fun for the employees. Chair trading is a pure exchange market. The two cases provide a clear case of models guiding conditional actions. Exchange markets work for chairs but not for desk locations.

## Models Without Lyapunov Functions

When we attempt to construct a Lyapunov function for a model and fail, we can still accumulate knowledge. Often we gain insight into why a model does not produce equilibria. In the Game of Life, some configurations attain equilibria, others do not. When one does produce an equilibrium, we might be able to write down a configuration-specific Lyapunov function. For example, any initial configuration that takes the form of a diagonal line will decrease by length 2 each period, as the two live cells at the end of the line die off and no new cells come to life. The configuration ends with all dead cells. For these configurations, the number of live cells would be a Lyapunov function. If we begin with another configuration, such as the R-pentomino, which produces a complex sequence of configurations, we cannot construct a Lyapunov function because the system does not go to equilibrium.

Our inability to construct a Lyapunov function does not imply that a model or system does not reach equilibrium. It could. Some systems go to equilibrium in every known case, yet no one has been able to construct a Lyapunov function. One famous example, the *half or triple plus one rule (HOTPO)* problem, also known as the Collatz conjecture, is deceptively simple. HOTPO starts with an integer. If the number is odd, we multiply by 3 and add 1. If the number is even, we divide by 2. The process stops when it reaches 1. If we start with the number 5 (which is odd), we triple it and add 1 to obtain 16. We divide 16 by 2 to get 8, divide 8 by 2 to get 4, and divide by 2 twice more to reach 1, the equilibrium. For every number up to  $2^{64}$ , HOTPO stops. Nevertheless, no one has proven whether HOTPO always goes to equilibrium. The mathematician Paul Erdos reportedly said, “Mathematics is not yet ripe for such problems.”<sup>5</sup> The inability of mathematicians to determine whether HOTPO goes to an equilibrium points to a more general lesson: models offer the possibility of proving results. We have no guarantee we can derive them. Often we write down a model only to find proving results difficult, if not impossible.

## Summary

In this chapter we have seen how Lyapunov functions help us to prove whether a system or model achieves equilibrium and how quickly it does. Even our failed attempts to construct a Lyapunov function can be of use. They can provide insights into the causes of complexity. Such was the case with exchange economies with externalities and the trading of desks. In neither case could we construct a global variable that would always decrease or increase. Thus, we had no guarantee that these processes would go to equilibrium.

When we think back to the uses of models—to reason, explain, design, communicate, act, predict, and explore—we find that Lyapunov functions can help with each. As just noted, Lyapunov functions help us to reason through why systems equilibrate. They can be used to design information systems such as Disney World’s sign-up times. We can use insights from the model to inform actions such as not allowing the trading of desks, to communicate how a system achieves equilibrium, to predict the time to reach equilibrium, and to explore, as we did when showing how cities self-organize.