# Lecture 4

*Lecturer: Giovanni Neglia* *Scribe: Miguel Romero, Melissa Sanabria*

**NOTE: The content of these notes has not been formally reviewed by the lecturer. It is recommended that they are read critically.**

## 1 Introduction

During the previous lecture we studied the dynamics of the bandwidth sharing over a network. For this analysis we took a non-rigorous approach when we assumed that the system must come to a halt (Lec. 3, Sec. 4.2), this reasoning led us to the conclusion that if the system stops it will do it at the optimum, but so far we don't know if the system really stops.

For this reason, in this lesson we will study the convergence of this dynamic system at the limit. In Section 2, we will study the behavior of the rates in some simple network configurations to observe if they eventually converge. We will find that the analysis of the rates in the network is not useful to determine if the system stops. Then, in Section 3, we will evaluate a different function from the system that will help us to determine whether the system stops and if it stops at the optimum. In Section 4, we will present the simulated annealing process that is a technique that will be utile to find the global optimum of a given function.

## 2 Does the system stop?

In the previous lecture we modeled the "Bandwith sharing over the Internet" problem as follows:

$$\begin{aligned} \underset{\mathbf{x},\mathbf{y}}{\text{maximize}} \quad & \sum_{r\in R} U_r(x_r) - \sum_{l\in E} M_l(y_l) \\ \text{subject to} \quad & y_l = \sum_{r|l\in r} x_r \\ & x_r \geq 0,\ \forall r \in R \end{aligned} \tag{1}$$

If we solve this problem by the Lagrange multipliers method, the optimal solution is:

$$U'_{\bar{r}}(x^*_{\bar{r}}) \begin{cases} = \sum_{l\in\bar{r}} M'_l(y^*_l) & \text{if } x^*_{\bar{r}} > 0, \\ \leq \sum_{l\in\bar{r}} M'_l(y^*_l) & \text{if } x^*_{\bar{r}} = 0. \end{cases} \tag{2}$$

This problem can be also solved using a distributed approach, where each network component has a role:

- The links measure congestion and transmit to the sources all the flow passing by them.
- The sources should adapt their rate over time according to the following equation:

$$\frac{dx_r}{dt} = k_r(U'_r(x_r(t)) - \sum_{l\in r} M'_l(y_l(t))) \tag{3}$$

We have concluded that if the elements of the network follow this dynamic and the system stops (the sources eventually stop changing their rate), then they stop at a point where the derivative (equation (3)) is equal to zero. Because otherwise, if the derivative is not zero, means that the system keeps changing.

If this derivative is equal to zero, then, from equation (3), we get:

$$U'_r(x_r(t)) = \sum_{l\in r} M'_l(y_l(t)) \tag{4}$$

We recognize, this is the same equation as the (strictly positive) optimal solution obtained by the lagrange multipliers method (equation (2)); but so far we have assumed that the system stops, let's observe with a few examples if the system actually stops.

## 2.1 First example: A single link with only one route

Let's first see the case of a single link with a single route, illustrated in Figure 1.

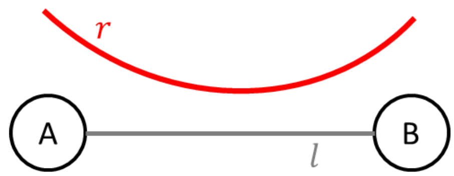



Figure 1: A single link with one route.

Assume that the equation of this network leads to the following:

$$\frac{dx_r}{dt} = k_r(1 - y_l) \tag{5}$$

But, as there is only one route, we have $y_l = \sum_{r|l \in r} x_r = x_r$, then:

$$\frac{dx_r}{dt} = k_r(1 - x_r) \tag{6}$$

The general solution of this differential equation is:

$$x_r(t) = 1 + ae^{-k_r t} \tag{7}$$

Imposing the initial condition $x_r(0) = 0$, we get that $a = -1$ and then we arrive to the solution:

$$x_r(t) = 1 - e^{-k_r t} \tag{8}$$

We observe in Figure 2 that this solution approaches to 1 exponentially fast, and in particular, it happens that the value of $k_r$ is the slope of the curve, it means that if $k_r$ is big, it grows fast.

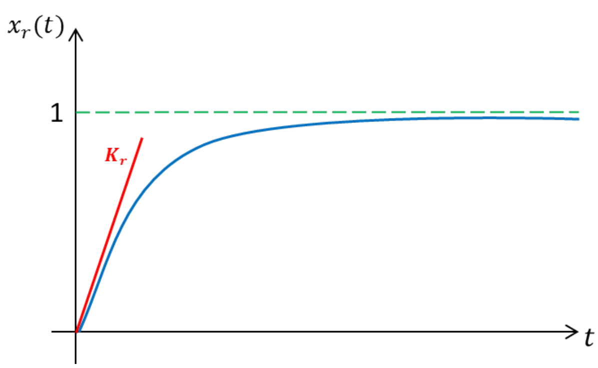



Figure 2: Dynamics of rate in route $r$.

Now, if we would like to solve our original maximization problem (equation (1)), we need the values of $U_r(x_r)$ and $M_l(y_l)$. From equation (3) and equation (5) we can deduce that:

$$U'_r(x_r) = 1 \implies U_r(x_r) = x_r \tag{9}$$

$$M_l'(y_l) = y_l \implies M_l(y_l) = \frac{{y_l}^2}{2} \tag{10}$$

Then, the maximization problem is:

$$\begin{aligned} \underset{\mathbf{x},\mathbf{y}}{\text{maximize}} \quad & x_r - \frac{{y_l}^2}{2} \\ \text{subject to} \quad & y_l = x_r \\ & x_r \geq 0, \ \forall r \in R \end{aligned} \tag{11}$$

Using the first constraint we can directly replace the value of $y_l$ by $x_r$.

$$\begin{aligned} \underset{\mathbf{x},\mathbf{y}}{\text{maximize}} \quad & x_r - \frac{{x_r}^2}{2} \\ \text{subject to} \quad & x_r \geq 0, \ \forall r \in R \end{aligned} \tag{12}$$

Using equation (9) and (10), in the solution of the optimization problem (equation (2)), we get:

$$1 \begin{cases} = x_r^* & \text{if } x_r^* > 0, \\ \leq x_r^* & \text{if } x_r^* = 0. \end{cases} \tag{13}$$

This equations tell us that either

$$\begin{cases} x_r^* = 1 \wedge x_r^* > 0 \implies x_r^* = 1 \\ x_r^* \geq 1 \wedge x_r^* = 0 \implies \textbf{contradiction} \end{cases}$$

Then, the optimal rate for this problem is $x_r^* = 1$. So, in this original problem, the source should transmit at rate 1.

We can observe, going back to the solution of the differential equation (equation (8)), that you go to this optimum $x_r^* = 1$, but the more you approach it the less is the speed at which you go there, then the convergence is asymptotic, you are decreasing your speed, but you are never stopping at the optimum.

## 2.2 Second example: Two routes sharing a link

We have seen until now that we converge to the optimum asymptotically, but although we are not stopping at the optimum, we are approaching it continuously, now we would like to see if this is always the case. Let's analyze a second example (illustrated in Figure 3), in which we observe the behavior of the rates in two routes using the same link.

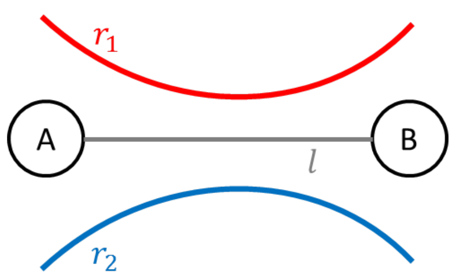



Figure 3: Two routes sharing a link.

Assume that these are the equations of the flows:

$$\frac{dx_{r_1}}{dt} = k_{r_1}\left(\frac{1}{x_{r_1}} - y_l\right) \tag{14}$$

$$\frac{dx_{r_2}}{dt} = k_{r_2}(\frac{2}{x_{r_2}} - y_l) \tag{15}$$

We can deduce the Utility functions of these flows as we have done for the previous example:

$$U'_{r_1}(x_{r_1}) = \frac{1}{x_{r_1}} \implies U_{r_1}(x_{r_1}) = \ln x_{r_1} \tag{16}$$

$$U'_{r_2}(x_{r_2}) = \frac{2}{x_{r_2}} \implies U_{r_2}(x_{r_2}) = 2\ln x_{r_2} \tag{17}$$

We can also deduce the Cost function of the link:

$$M'_l(y_l) = y \implies M_l(y_l) = \frac{y^2}{2} \tag{18}$$

#### 2.2.1 $r_1$ alone

First, let's assume that at the beginning $r_1$ is alone, then developing a similar procedure as for the previous example, the condition would be:

$$\frac{1}{x^*_{r_1}} \begin{cases} = x^*_{r_1} & \text{if } x^*_{r_1} > 0, \\ \leq x^*_{r_1} & \text{if } x^*_{r_1} = 0. \end{cases} \tag{19}$$

This equations tells us that either $\begin{cases} {x^*_{r_1}}^2 = 1 \land x^*_{r_1} > 0 \implies x^*_{r_1} = 1 \\ {x^*_{r_1}}^2 \geq 1 \land x^*_{r_1} = 0 \implies \textbf{contradiction} \end{cases}$

Then, the optimal value is $x^*_{r_1} = 1$.

#### 2.2.2 $r_1$ and $r_2$ together

Now, let's look at the case where $r_1$ and $r_2$ are together. We can deduce the Utility function of these flows as follows, for $r_1$:

$$U'_{r_1}(x^*_{r_1}) \begin{cases} = M'_l(y^*_l) & \text{if } x^*_{r_1} > 0, \\ \leq M'_l(y^*_l) & \text{if } x^*_{r_1} = 0. \end{cases} \tag{20}$$

As we know that $U'_{r_1}(x_{r_1}) = \frac{1}{x_{r_1}}$ and $M'_l(y_l) = y_l = \sum\limits_{r|l\in r} x_r = x_{r_1} + x_{r_2}$, then, replacing:

$$\frac{1}{x_{r_1}} \begin{cases} = x^*_{r_1} + x^*_{r_2} & \text{if } x^*_{r_1} > 0, \\ \leq x^*_{r_1} + x^*_{r_2} & \text{if } x^*_{r_1} = 0. \end{cases} \tag{21}$$

Similarly for $r_2$:

$$\frac{2}{x_{r_2}} \begin{cases} = x^*_{r_1} + x^*_{r_2} & \text{if } x^*_{r_2} > 0, \\ \leq x^*_{r_1} + x^*_{r_2} & \text{if } x^*_{r_2} = 0. \end{cases} \tag{22}$$

Let's suppose that $x^*_{r_1} > 0$ and $x^*_{r_2} > 0$, then we have a system of two equations:

$$\frac{1}{x^*_{r_1}} = x^*_{r_1} + x^*_{r_2} \tag{23}$$

$$\frac{2}{x^*_{r_2}} = x^*_{r_1} + x^*_{r_2} \tag{24}$$

Subtracting equations (23) and (24):

$$\frac{1}{x^*_{r_1}} - \frac{2}{x^*_{r_2}} = 0 \implies x^*_{r_2} = 2x^*_{r_1} \tag{25}$$

Using equation (25) in equation (23), we get:

$$3{x^*_{r_1}}^2 = 1 \implies x^*_{r_1} = \frac{1}{\sqrt{3}} \tag{26}$$

Replacing equation (26) in equation (25):

$$x^*_{r_2} = \frac{2}{\sqrt{3}} \tag{27}$$

This result is coherent, because as they are sharing the same link and the utility of $r_2$ is higher than the utility of $r_1$, is reasonable that at the end (in the optimal case) the rate of $r_2$ is higher than the rate of $r_1$.

Now with this information, we can try to see what can be the dynamics of the rates.

### 2.2.3 Dynamics $r_1$ alone

Let's start trying to deduce the dynamics of $x_{r_1}$ for the first case. When $r_1$ is alone, it will converge to 1, we don't know the exact shape, but it can be similar to the one presented in Figure 4.

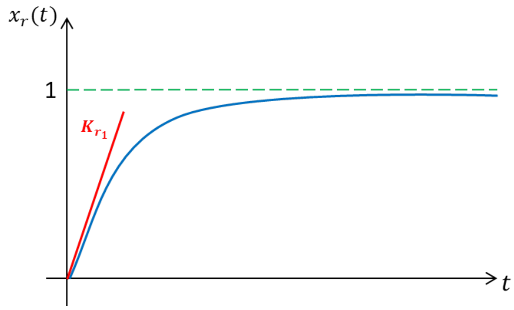



Figure 4: Behavior of $r_1$ alone.

In particular, the speed at which it approaches 1 will be related to $k_{r_1}$.

### 2.2.4 Dynamics $r_1$ and $r_2$ together

Now, let's take a look to the behavior of $x_{r_1}$ and $x_{r_2}$ when $r_1$ and $r_2$ are sharing the link. In this case, it is more complex to know what will happen, because $x_{r_1}$ has its target in $\frac{1}{\sqrt{3}}$ and the target of $x_{r_2}$ is $\frac{2}{\sqrt{3}}$. Again, the speed at which they approach to their targeted value will depend on the parameter $k_{r_1}$ and $k_{r_2}$.

Let's assume that $k_{r_2}$ is much smaller than $k_{r_1}$:

$$k_{r_2} << k_{r_1}$$

so that $x_{r_2}$ changes but very slowly.

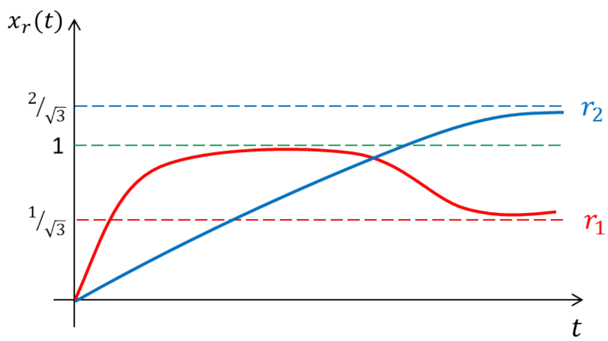



Figure 5: Behavior of $r_1$ and $r_1$ together.

So, as illustrated in Figure 5, at the beginning the influence of $r_2$ will be so low that $r_1$ will not notice it, and then $r_1$ will start to increase its rate, $x_{r_1}$ will tend to go to 1, as if it were alone in the link.

Meanwhile $r_2$ will be increasing its rate slowly. At one point $x_{r_2}$ will start becoming significant, and the more $x_{r_2}$ grows, the more $x_{r_1}$ will start perceiving the congestion, until $x_{r_1}$ has to come back to its final (optimal) condition.

We can see that actually, even if there are only two flows sharing the link, it is possible that at the beginning one of them overshoots its target to perhaps come back later, and these dynamics could be even more complex as we will see with the following examples.

## 2.3 Third example: More complex networks

We have seen that for a simple network with only one link and two routes sharing the link, the equations and the behavior of the flows start to be complex, now, let's take a look to the configuration shown in Figure 6, where $r_1$ and $r_2$ are competing for link $l_2$, and $r_2$ and $r_3$ are competing for link $l_3$.

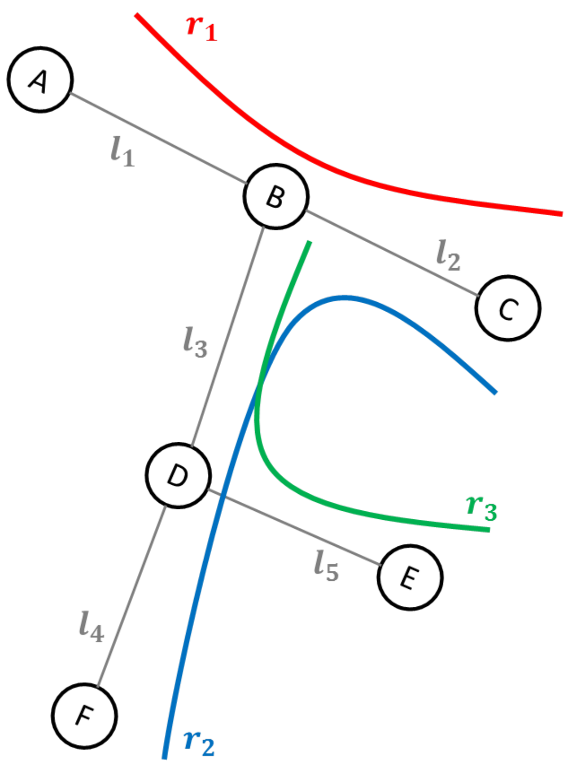



Figure 6: Example of a network configuration with 3 routes over 5 links.

Let's say that the utilities of the sources $r_1$, $r_2$, and $r_3$ are the following:

$$
\begin{aligned}
U_{r_1}(x_{r_1}) &= \ln x_{r_1} \\
U_{r_2}(x_{r_2}) &= 2 \ln x_{r_2} \\
U_{r_3}(x_{r_3}) &= 10 \ln x_{r_3}
\end{aligned}
\tag{28}
$$

Now, assume that $k_{r_1} >> k_{r_2} >> k_{r_3}$. In Figure 7 we observe the possible behavior of flow $x_{r_1}$: At the beginning $x_{r_1}$ will grow pointing to 1, when $x_{r_2}$ starts to be significant, $x_{r_1}$ will start to decrease towards $\frac{1}{\sqrt{3}}$, but then after some time, $x_{r_3}$ will be high enough to compete with $x_{r_2}$ on link $l_3$, and because it has larger utility, it will probably reduce the rate of $r_2$. Then $r_1$ will have more space on link $l_2$, and will increase its rate again approaching to 1, and the dynamics continue.

Then we can build a more complex network configuration like the one proposed in Figure 8, where it is possible that a behavior like the following occurs: First $x_{r_1}$ will grow towards 1, then $x_{r_2}$ starts to be

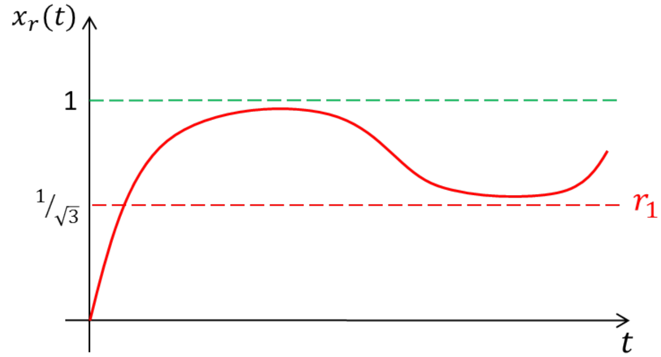



Figure 7: Flow $x_{r_1}$ over time.

significant, and $x_{r_1}$ decrease approaching its goal value. After some time, $x_{r_3}$ will be noticed by $x_{r_2}$ which will reduce its own rate, causing that $x_{r_1}$ increases again. At some point, $x_{r_4}$ will be perceived by $x_{r_3}$ which will reduce its own rate, this will yield the increasing of $x_{r_2}$ and consequently the reduction of $x_{r_1}$. In general, the dynamics will continue running and the rates can oscillate going up and down, now it's not clear anymore even if they asymptotically converge to an equilibrium point.

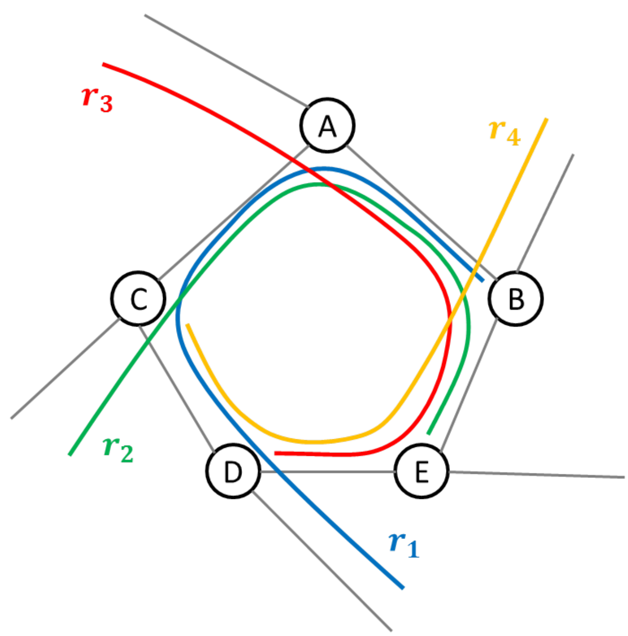



Figure 8: Network configuration with 10 links and 4 flows.

# 3 Does the system converge?

We wanted to find something like the following:

$$\lim_{t\to\infty} x_r(t) = x_r^*$$

In general: given a system of equations:

$$\dot{x}_{r_1}(t) = f_1(x_{r_1}, x_{r_2})$$

$$\dot{x}_{r_2}(t) = f_2(x_{r_1}, x_{r_2})$$

We would like to prove that:

$$\dot{x}_{r_1}(t) \xrightarrow[t\to\infty]{} x^*_{r_1}$$

$$\dot{x}_{r_2}(t) \xrightarrow[t\to\infty]{} x^*_{r_2}$$

One possibility to prove this is to observe the variable $x_{r_1}$ over time, and show that it is each time closer to $x^*_{r_1}$, then do the same for $x^*_{r_2}$. Once you have proven that each of the variables $x_r$ in the system converge to $x^*_r$ over time, the prove is finished.

What we have seen is that it is not possible to use this proof method, because from the study of the flows in the previous section we have observed the following:

- Probably the system will not converge in finite time.
- It is possible that we are in a chaotic system. We are not sure if the system will eventually converge.

Instead of analysing the rates, because as we have seen in the previous section they can oscilate, and it is not clear how this can lead us to some convergence, we want to use another function that helps us.

We can think here on the function we used in the maximization problem:

$$V(t) = \sum_{r\in R} U_r(x_r(t)) - \sum_{l\in E} M_l(y_l) \tag{29}$$

We can wonder, is it possible that although each individual rate does not go closer and closer over time to the optimal solution, this global function (equation (29)) keeps going closer and closer to the optimal solution?

To answer this, we have to solve two questions:

1. Does $V(t)$ always increase unless it reaches the optimum?

$$\frac{dV(t)}{dt} > 0,\ \forall t\ \ unless\ V(t) = \max V(x_r)\ ?$$

2. Does this function achieve the point of maximum?

$$\lim_{t\to\infty} V(t) = \max V(x_r)\ ?$$

If the answers to this two questions are Yes, imply that:

$$\lim_{t\to\infty} x_r(t) = x^*_r$$

Because $x^*_r$ is the only point where the function $V(t)$ is maximum, and if $V(t)$ is always increasing, then once it reaches the maximum, it cannot continue changing/increasing more over time.

We will use the following additional hypothesis:

$$U'_r(0) = +\infty,\ \forall r \in R$$

If this is true, then we know that $x^*_r > 0$, and it means that we can only consider the simplest equations (equality equations) in the condition of equation (2).

## 3.1 Is $V(t)$ always increasing?

To know if the function $V(t)$ is increasing with time, we compute the derivative (using the chain rule):

$$\frac{dV(t)}{dt} = \frac{dV(x_{r_1}(t), x_{r_2}(t), ..., x_{r_R}(t))}{dt} = \sum_{r\in R} \frac{\partial V}{\partial x_r} \cdot \frac{dx_r}{dt} \tag{30}$$

Computing the partial derivative $\partial V/\partial x_r$:

$$\frac{\partial V}{\partial x_r} = (U_r'(x_r(t)) - \sum_{l\in r} M_l'(y_l(t))) \tag{31}$$

And we already had the result of $dx_r/dt$ in equation (3). Then replacing (3) and (31) in (30):

$$\frac{dV(t)}{dt} = \sum_{r\in R} (U_r'(x_r(t)) - \sum_{l\in r} M_l'(y_l(t)))^2 \cdot (k_r) \tag{32}$$

We have the squared term of $(U_r'(x_r(t)) - \sum_{l\in r} M_l'(y_l(t)))$, multiplied by a positive term $(k_r)$, and then we have the sum over these positive terms, then we can conclude that the value of $dV(t)/dt$ is always positive.

$$\frac{dV(t)}{dt} \geq 0, \ \forall t \tag{33}$$

If we check the point where the derivative is equal to zero:

$$\frac{dV(t)}{dt} = 0 \implies U_r'(x_r(t)) = \sum_{l\in r} M_l'(y_l(t)) \tag{34}$$

We can remember from equation (2) that this is the same optimal solution obtained by the method of lagrange multipliers. It implies a much stronger condition, this function is always increasing unless we arrive to the optimum:

$$\frac{dV(t)}{dt} > 0, \ \forall t \ \ unless \ x_r(t) = x_r^*(t) \tag{35}$$

## 3.2 Does $V(t)$ go towards $V(x^*)$?

Now the question that arises is, does $V(t)$ go towards $V(x^*)$?

Because it could be that the function is increasing but each time slower that it approaches assymptotically to the optimum, and never reaches it.

So, what does it mean that the function $V(t)$ never reaches the maximum? Let's try to illustrate the space of $x_r$ in Figure 9.

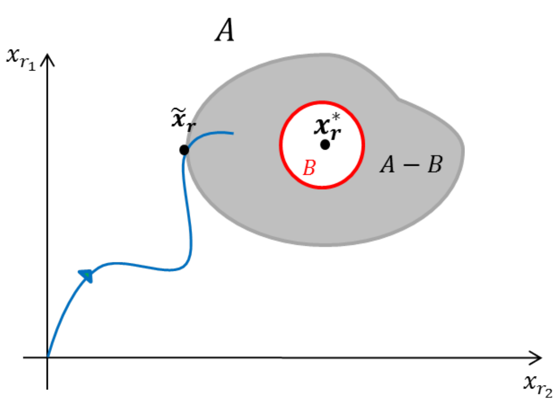



Figure 9: Evolution (over time) of $x_r$ in the space, plot for two routes, $R = r_1, r_2$.

This Figure plots the evolution over time of the values of $x_r$, then we start from $x_r = 0$, and we target to the optimum (it can be targeted in a strange way as shown in the figure), this is the curve over time, and let's suppose it approaches assymptotically to $x_r^*$.

At one instant we will be at point $\widetilde{x}_r$, now let's define the set $A$ to be:

$$A = \{x_r \ | \ V(x_r) \geq V(\widetilde{x}_r)\} \tag{36}$$

The set $A$ will include $\widetilde{x}_r$ and $x_r^*$ (the optimum), as shown in Figure 6. This set is compact (closed and bounded).

Let's assume now that we are approaching $x_r^*$ asymptotically, this means that we will never reach it, and then there will exist a ball $B$ around $x_r^*$ such that $x_r(t)$ will never be there at any time.

$$\text{If } x_r(t) \not\to x_r^* \implies \exists\ B \ | \ x_r(t) \notin B \ \forall t \tag{37}$$

Now, if we look at the set $A - B$, we find that in this set, $dV/dt$ is strictly positive, because we already found that this derivative is always positive, except in $x_r^*$ and this point does not belong to the set $A - B$.

$$\frac{dV}{dt} > 0 \ \text{ over } A - B$$

Then, the function $dV/dt$ happens to be continuous, and we have seen that set $A$ is a compact set.

**Claim 1.** *A continuous function over a compact set implies that the function has a maximum and a minimum on the set.*

This is telling us that $V(t)$ moves inside $A - B$ with a strictly positive speed ($a$).

$$\frac{dV}{dt} \text{ is continuous over a compact set } \implies \exists \min \frac{dV}{dt} = a > 0$$

This implies that $V(t)$ cannot stay forever out of set $B$. In other words, it cannot be that $V(t)$ is always increasing but it never goes inside set $B$. Then we arrived to a contradiction, because we supposed that $V(t)$ could have stayed always inside the set A-B and we found that it does not happen.

Our proof is finished. Then we can conclude that:

**Remark 1.** *We started from an optimization problem, and we have shown that the dynamics we have defined converge to the optimum.*

**Remark 2.** *From another point of view, if you don't have the optimization problem, but you have the dynamics of the system, and you wonder if they converge or not. We can prove it by defining a real function that always increases. This function is usually called "potential function" or "Lyapunov function".*

# 4 Simulated Annealing — GIBBS Sampling

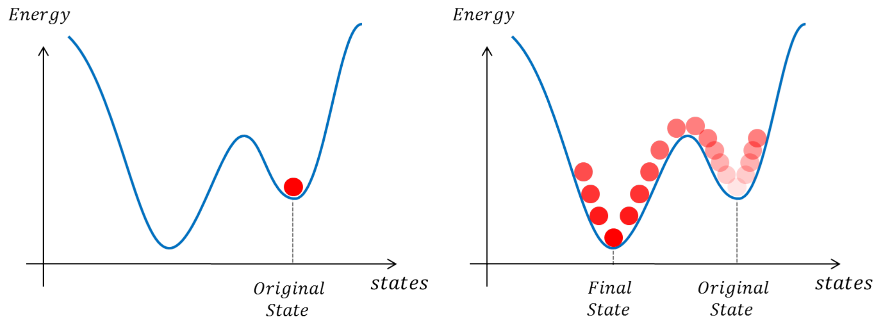
Annealing is a heat process whereby a metal is heated to a specific temperature and then allowed to cool down slowly. This softens the metal which means it can be cut and shaped more easily.

The idea of this procedure is to heat the metal in order to cause the oscillation of the atoms which will potentially break the disordered pieces of the metal. The temperature is slowly reduced to make the atoms move until they reach some ordered state, which is the minimum energy configuration.

If this configuration has the minimum energy configuration, the question is why the metal by itself does not go directly to this configuration?

It happens because the metal presents an analog situation to the one shown in Figure 10. This figure represents a system with different states, which can be all the different positions of the atoms in the metal. Let's assume that there is a ball that arrives to this curve in a point (state), and you know there is another state in the system that has less energy, but in order to reach that point, the ball need to pass first through a space where the energy is higher.

If you can increase the kinetic energy of the ball, at one point, the ball will be able to pass by the configuration of higher energy, and then will reach the minimum energy configuration; meanwhile you can reduce gradually the energy until the ball is again motionless, but now in the global minima.




(a) Original State of the system.

(b) Trajectory of the ball from the initial state after aplying some kinetic force until it arrives to the final state.

Figure 10: Energy of the system for the different states.

Now, we can formalize the definition of the problem. Let's assume we have a set of states $S$, and a number $N$ of agents. Each agent $i$ has some internal variable $a_i \in A_i$ where $A_i$ is a finite set. We also define the state as $S(a_1, a_2, ..., a_N)$. Then, if you have a finite number of agents ($N$), and they have a finite number of possibilities ($\mid A_i \mid$), the space is finite (and not continuous as the problems seen in the previous sections).

Let's see an example. Suppose that action $a_i \in \{0, 1\}$, and the state of the system is characterized by Equation 38 which describe how many nodes have done a particular choice, then a possible plot of this example is presented in Figure 11.

$$S(a_1, a_2, ..., a_N) = \sum_{i=1}^{N} \mathbb{1}(a_i = 1) \tag{38}$$

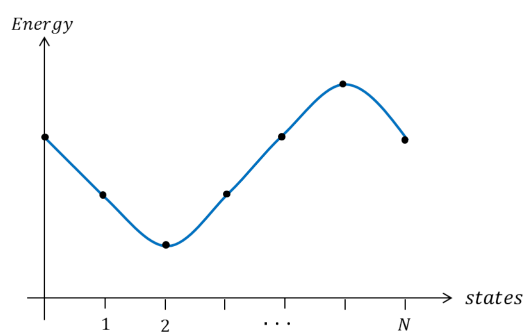



Figure 11: State of the system.

To solve this simple example, we want to minimize the energy $\xi$ that is a function of state $S$ (Equation 39), the goal is to find the state for which the energy is minimal.

$$\underset{\{a_i\}}{\text{minimize}} \quad \xi(S(a_1, ..., a_N)) \tag{39}$$

There are some strong differences with the problems we solved before. Now, the solution space is finite and we do not rely on assumptions about the shape of the function (concave, convex, increasing, etc).

## 4.1 Neighborhood Relation

One way to solve the problem described before is to have a neighborhood relation, meaning that from one state you can consider some other state. This relation is shown in Figure 12 where we can move from state $S_i$ to state $S_j$, but cannot move directly from state $S_i$ to state $S_k$.

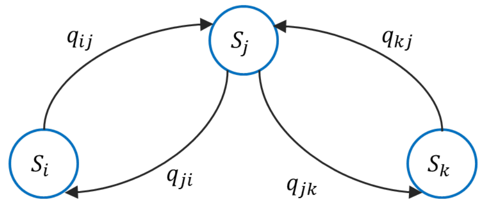



Figure 12: Representation of Neighborhood relation.

First, let's describe some definitions.

- $q_{ij}$ is the probability of state $S_i$ to consider state $S_j$
- $\xi(S_j) = \xi_j$ is the energy of state $S_j$
- $\alpha_{ij}$ is the probability to move from state $S_i$ to state $S_j$ given that you have considered state $S_j$

## 4.2 Modelling the problem as a Markov Chain

Now, let's define some rules:

1. At every step you can consider one of your neighbors with probability $q_{ij}$.
   We will assume that $q_{ij} = q_{ji}$.
2. Once you decide to look into node $j$, you can compare its energy with your energy according to Equation 40, where we can observe that if the difference is small, the probability to move to state $S_j$ is high.

$$\alpha_{ij} = \begin{cases} 1 & \text{if } \xi_j \leq \xi_i, \\ e^{-\frac{\xi_j - \xi_i}{T}} & \text{if } \xi_j > \xi_i. \end{cases} \tag{40}$$

From (40) we can observe that if the energy of state $S_j$ is smaller than the energy in the current state, I will always move there. On the other hand, if the energy of state $S_j$ is greater than my current energy, I will move depending on the difference of the energies. If the difference is small, then I would go there with a high probability but if the difference is very large, it will be very unlikely that I move to state $S_j$.

After all this, we have found a Markov chain (Figure 13), where the probability to go from $S_i$ to $S_j$ is $q_{ij} \cdot \alpha_{ij}$.

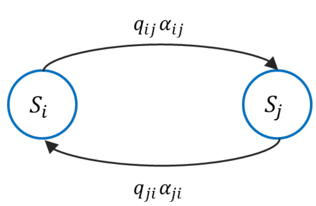



Figure 13: Markov chain.

If this Markov Chain is reversible, we know that it will converge to the stationary distribution described in (41), meaning that if we let this markov chain run during a long time, we know that with some probability $\pi_i$ it will be at state $S_i$, if we assume the Markov Chain is reversible, then this probability is:

$$\pi_i = \frac{e^{-\frac{\xi_i}{T}}}{\sum_j e^{-\frac{\xi_j}{T}}} \tag{41}$$

To check if (41) is really the stationary distribution of our problem, we have to proof the reversibility property of this Markov Chain. It means that if we take any two neighbors $S_i$ and $S_j$ the rate at which I go from $S_i$ to $S_j$, $(\pi_i \cdot q_{ij} \cdot \alpha_{ij})$, and the rate at which I go from $S_j$ to $S_i$, $(\pi_j \cdot q_{ji} \cdot \alpha_{ji})$, should be equal:

$$\pi_i \cdot q_{ij} \cdot \alpha_{ij} = \pi_j \cdot q_{ji} \cdot \alpha_{ji} \tag{42}$$

As $q_{ij} = q_{ji}$ we have:

$$\pi_i \cdot \alpha_{ij} = \pi_j \cdot \alpha_{ji} \tag{43}$$

Replacing (41) in (43), and simplifying, we obtain:

$$e^{-\frac{\xi_i}{T}} \cdot \alpha_{ij} = e^{-\frac{\xi_j}{T}} \cdot \alpha_{ji} \tag{44}$$

Now, if we assume $\xi_j > \xi_i$, we can replace $\alpha_{ij}$ and $\alpha_{ji}$ using equation (40):

$$\begin{aligned} e^{-\frac{\xi_i}{T}} \cdot e^{-\frac{\xi_j - \xi_i}{T}} &= e^{-\frac{\xi_j}{T}} \cdot 1 \\ e^{-\frac{\xi_i}{T} - \frac{\xi_j - \xi_i}{T}} &= e^{-\frac{\xi_j}{T}} \\ e^{\frac{-\xi_i - \xi_j + \xi_i}{T}} &= e^{-\frac{\xi_j}{T}} \\ e^{\frac{-\xi_j}{T}} &= e^{-\frac{\xi_j}{T}} \end{aligned} \tag{45}$$

So, we have found an equality, proving that this Markov chain is reversible and now we can use directly the stationary distribution as the solution of our problem.

## 4.3 Reducing the temperature

First, let's analyze how could be the graph of the function $\pi_i$ (equation 41).

$$\pi_i = \frac{e^{-\frac{\xi_i}{T}}}{\sum_j e^{-\frac{\xi_j}{T}}}$$

We can observe from $\pi_i$, that when $\xi_i$ is big, $e^{-\frac{\xi_i}{T}}$ is close to 0, and then $\pi_i$ will be small. When $\xi_i$ is small, $e^{-\frac{\xi_i}{T}}$ is close to 1, and then $\pi_i$ will be big.

So we can now plot the probability of being in a given state $i$, in Figure 14.

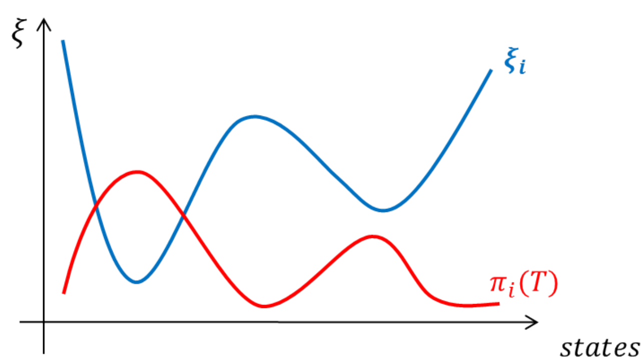



Figure 14: Stationary distribution (red) compared with the curve of the energy of the states (blue).

Let's now analyze what happens when we start decreasing the temperature $T$. If the temperature decreases, the numerator and the denominator of $\pi_i$ decrease, but which one decreases faster? If we compute the ratio between $\pi_i$ and $\pi_j$, we get:

$$\frac{\pi_i}{\pi_j} = \frac{\frac{e^{-\frac{\xi_i}{T}}}{\sum_k e^{-\frac{\xi_k}{T}}}}{\frac{e^{-\frac{\xi_j}{T}}}{\sum_k e^{-\frac{\xi_k}{T}}}} = \frac{e^{-\frac{\xi_i}{T}}}{e^{-\frac{\xi_j}{T}}} = e^{+\frac{\xi_j - \xi_i}{T}}$$

If $\xi_i$ is smaller than $\xi_j$, then the difference $(\xi_j - \xi_i)$ is positive. If $T$ goes to zero, this ratio goes to infinity. If now we say that $\xi_i$ is greater than $\xi_j$, and $T$ goes to zero, this ratio will tend to zero.

This means for $\pi_i$, that when $T$ becomes smaller, the numerator will decrease, but will decrease much faster in the states with high energy. Then, the states with higher energy will become less and less likely and the states with low energy will become more and more likely, we plotted this in Figure 15.

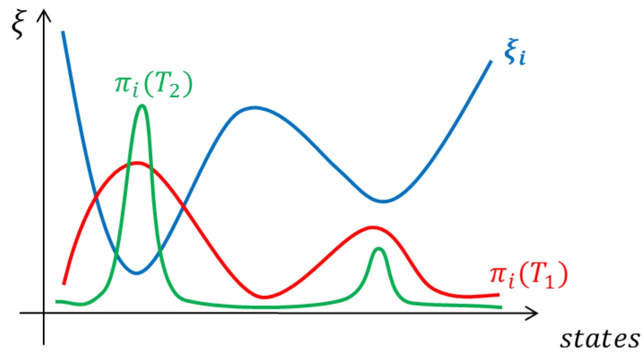



Figure 15: Stationary distribution for a temperature $T_1$ (red) compared with the Stationary distribution for a temperature $T_2 < T_1$ (green) and the curve of the energy of the states (blue).

Let's define now $H$ as the set of states with minimum energy. Then, we can prove that if you leave the temperature go to zero:

$$\lim_{T \to 0} \pi_i(T) = \begin{cases} \frac{1}{|H|} \text{ if } i \in H \\ 0 \text{ otherwise} \end{cases} \tag{46}$$

If the temperature reaches zero, you will be sure that your system will be in one of the global minima of the system. If you have only one global minima $\mid H \mid = 1$, the probability to be in this global minima will be equal to 1.

But we cannot decrease the temperature very fast. Because, from equation (40), if $T \to 0$ the probability to move from another state with higher energy will be zero, and then we will possibly be in a local minima.

$$\alpha_{ij} = \begin{cases} 1 & \text{if } \xi_j \le \xi_i \\ e^{-\frac{\xi_j - \xi_i}{T}} & \text{if } \xi_j > \xi_i \end{cases} \xrightarrow[T \to 0]{} \alpha_{ij} = \begin{cases} 1 & \text{if } \xi_j \le \xi_i \\ 0 & \text{if } \xi_j > \xi_i \end{cases}$$

We would like to have at the beginning a temperature high enough so that we can reach the stationary distribution, and then reduce the temperature until it reaches zero.

Now we have to combine the two things:

- We have to run the system a long time in order to arrive to the stationary distribution.
- And then we have to reduce the temperature.

We can think of reducing the temperature step by step. You start from an inicial temperature at step 0, and then you divide this temperature by $\ln(1 + k)$, where $k$ is the step.

This reduction of the temperature in function of the step $k$ is defined by equation (47), and is represented in Figure 16.

$$T(k) = \frac{T_0}{ln(1 + k)} \tag{47}$$

Then, you are guaranteed that you will assymptotically be in one of the global minima, if you execute the Algorithm 1.

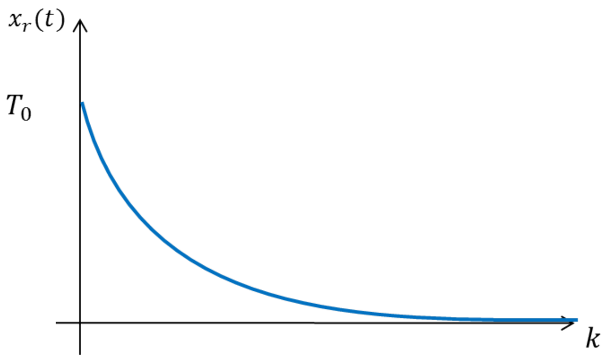



Figure 16: Decreasing of the temperature in function of the step $k$.

**Algorithm 1** Simmulated Annealing.

**Require:**
$S$ : Set of states
$T_0$ : Initial temperature
$\xi$ : Energy of the states
$\epsilon$ : tolerance
**Ensure:**
$i$ : State of global minima
**procedure** SimmulAnnealing
    $i \leftarrow$ any state of $S$
    $T \leftarrow T_0$
    $k \leftarrow 1$
    **while** $T > 0 + \epsilon$ **do**
        $j \leftarrow$ select a neighbor of $i$ with probability $p_{ij}$
        **if** $\xi_i > \xi_j$ **then**
            $i \leftarrow j$
        **else**
            $i \leftarrow j$ with probability $e^{-\frac{\xi_j - \xi_i}{T}}$
        $T \leftarrow T_0 / \ln(1+k)$
        $k \leftarrow k+1$
    **return** $i$

# References

[1] Frank Kelly and Elena Yudovina, Stoch astic Networks. *Cambridge Press*, 2014.