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

1 Introduction

In the previous lesson, we looked into the example of an electrical network and showed that a simple local policy for an agent, in our case an electron, can lead to the optimization of a large distributed problem which is the minimization of energy loss in the network. In this lesson, we consider a road traffic model to demonstrate that choosing local policies is not trivial and how potential wrong choices can lead to poor system behaviour. Moreover we solve the more general problem which arises from our road network and present how local policies in such cases should be correctly chosen.

2 Problem Definition

We consider an oriented graph (e.g. the road-traffic graph of a city) like in Figure 1. From now on we look at this problem as a routing problem.
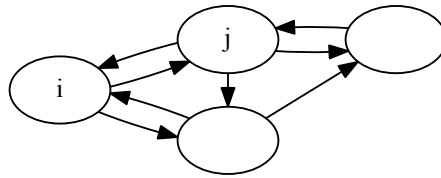

Figure 1: Oriented graph of a city

We have a set of directed links E which are the connections between the nodes of the graph and a set of *source-destination* pairs S, where we have $s \in S$ such that $s = (i, j)$ represents the agents wanting to travel from node i to j. Moreover we denote by R the set of all routes.

For each $s \in S$ we define $R(s)$ to be the set of routes connecting the *source-destination* pair and f_s to be the amount of traffic to route on s. Also $s(r)$ is the *source-destination* pair connected by $r \in R$. Since a given f_s can be split over various routes, we denote by x_r the amount of traffic of $s(r)$ that takes route r i.e. $\sum_{r \in R(s)} x_r = f_s$, $\forall s$ with $x_r \geq 0$.

Finally, we define y_l to be the amount of traffic on link $l \in E$ due to all routes passing by l, i.e. $y_l = \sum_{r|l \in r} x_r$. On each link l there exists a delay per unit of traffic which is a function of the traffic on that link that we denote by $D_l(y_l)$. The delay function is assumed to be convex, increasing and its second derivative to exist.

Thus the problem consists of minimizing the overall delay experienced by the agents.

The optimization problem can be modeled as follows:

$$\text{minimize} \quad \sum_{l \in E} y_l D_l(y_l)$$

$$\text{subject to} \quad \sum_{r \in R(s)} x_r = f_s, \ \forall s \in S \tag{1}$$

$$y_l = \sum_{r | l \in r} x_r, \ \forall l \in E$$

$$x_r \geq 0, \ \forall r \in R$$

This problem happens to be convex.

Proof: We first observe that the constraints identify a convex set $\{x_r \geq 0\}$.

We know that D_l is increasing and convex. This means that for all y_l we have $D_l'(y_l) \geq 0$ and $D_l''(y_l) \geq 0$.

We know that $y_l D_l(y_l)$ is convex iff its second derivative is nonnegative.

We start by computing the first derivative. $(y_l D_l(y_l))' = D_l(y_l) + y_l D_l'(y_l)$ where $D_l(y_l) \geq 0$ and $y_l D_l'(y_l) \geq 0$. Thus $(y_l D_l(y_l))'$ is nonnegative.

Then we compute the second derivative. $(y_l D_l(y_l))'' = D_l'(y_l) + y_l D_l''(y_l) + D_l'(y_l) = 2D_l'(y_l) + y_l D_l''(y_l)$ where $2D_l'(y_l) \geq 0$ and $y_l D_l''(y_l) \geq 0$. Thus $(y_l D_l(y_l))''$ is nonnegative.

Hence $y_l D_l(y_l)$ is convex and it implies that $\sum_l y_l D_l(y_l)$ is also convex and thus this problem is convex. $\square$

3 Naive solution

One naive solution for this problem would be that each agent tries to minimize its own delay, and from this we could hope that the total delay would be minimized. Therefore we are going to check if this solution is indeed the optimal one and, if it is not so, how far off we are from the optimal.

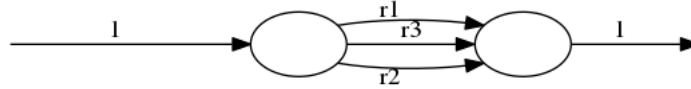## 3.1 A first example: parallel links with linear delay functions

Figure 2: Our specific network

With this specific example each different edge has a specific delay equal to $D_i = \alpha_i \times y_i$.

In this setting, if every agent minimizes its own delay, we can expect that at the equibrium all the routes have the same delay. Indeed, if it were not so, some of the agents will find more convenient to move from a route with higher delay to a route with lower delay.

We can now check if this equilibrium corresponds to the solution minimizing the total delay. The total delay is $\sum_{i=1}^{3}(\alpha_i \times y_i) \times y_i = \sum_{i=1}^{3} \alpha_i \times y_i^2$, and we can see that this function is the same as the model we studied in the first lecture.

We can then compare our network to an electrical network with three different resistances in parallel, and from our previous lecture we know that at the optimum, the currents are such that $\Delta V = I_1 R_1 = I_2 R_2 = I_3 R_3$ with $I_i = y_i$ and $R_i = \alpha_i$, so at the optimal solution y_i^* we have then $\alpha_1 \times y_1^* = \alpha_2 \times y_2^* = \alpha_3 \times y_3^*$ which is $D_1(y_1^*) = D_2(y_2^*) = D_3(y_3^*)$.

The conclusion is then that, at least in this setting, if each agent selects the route with minimum delay, then the system converges to a configuration with minimum global delay.
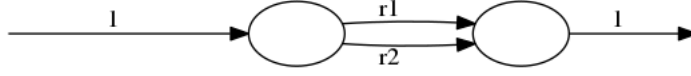
3.2 Pigou's example

Figure 3: Our general network

We are now going to extend the previous example with delay functions that are not linear, we have then $D_1(y_1) = 1$ and $D_2(y_2) = y_2^n$, because this function relates the most to how delay affects routing in a real world scenario. This example is known as the Pigou's example.

The comparison between the two delay function gives us the following graphic.
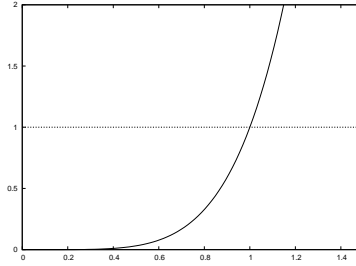

Figure 4: Our general network

We now put our naive solution in this example with each person minimizing his or her delay. We get $y_1^1 = 0$ and $y_2^1 = 1$ because the second link is always the most attractive with regard to the delay. The total delay is then $D_t(1) = 1 \times 0 + 1^n \times 1 = 1$.

We now solve the problem. Consider a general splitting of the traffic between the two links and denote by α the amount of traffic on link 2. The traffic on link 1 is then $1 - \alpha$ and the total delay is $D_t(\alpha) = (1 - \alpha) \times 1 \times \alpha \times \alpha^n$. We know that to find the minimum we look for:

$$D_t'(\alpha) = 0 \iff -1 + (n+1) \times \alpha^n \iff \alpha^* = \frac{1}{\sqrt[n]{n+1}}$$

Thus we know that the optimal quantity of traffic is α^*. Now we will check if the delay we got from the naive solution is optimal or not.

The optimal delay is:

$$D_t(\alpha^*) = 1 - \alpha^* + \alpha^{*^{n+1}} = 1 - \frac{1}{\sqrt[n]{n+1}} + \frac{1}{\sqrt[n]{n+1}} * \frac{1}{n+1} = 1 - \frac{1}{\sqrt[n]{n+1}} * \frac{n}{n+1}$$

Thus we can deduce that for every $n > 0$ the delay is strictly less than 1. Therefore the naive local rule does not lead to minimize the total delay. Nevertheless, if the corresponding total delay is close enough to the optimal one, it would still be acceptable.

For this we compute the price of anarchy:

$$\frac{D_t(\alpha^1)}{D_t(\alpha^*)} = \frac{1}{1 - \frac{1}{\sqrt[n]{n+1}} * \frac{n}{n+1}} \xrightarrow[n \to \infty]{} +\infty$$

Thus the naive solution gives us a delay that is arbitrarily bigger than the optimal one so our solution is unfortunately pretty poor.

After having ruled out the naive local rule to select the route with minimum delay, we move to study directly the original problem.

4 Solving Problem (1) by the Lagrange multipliers method

Observe that the feasible set is compact (closed and bounded). In fact the constraints define a bounded polyhedral set because each variable x_r is bounded:

$$0 \leq x_r \leq \max_{s \in S} f_s \quad \forall r \in R \tag{2}$$

Because the feasible set is compact and the objective function is continuous, then the problem admits a global minimum.

Moreover, all the hypotheses of theorem 1 from lesson 2a are satisfied and then we can find a point of global minimum using the Lagrange multiplier method. In particular we consider as polyhedral set $X = \{x_r \geq 0, \forall r \in R\}$ and we define two set of multipliers:

1. one for each source-destination pair, λ_s, $\forall s \in S$ and

2. one for each link, ν_l, $\forall l \in E$.

Let $\mathbf{z}^* = \begin{pmatrix} \mathbf{x}^* \\ \mathbf{y}^* \end{pmatrix}$ denote a global minimizer, then, because of theorem 1, there exist λ_s^*, $\forall s \in S$ and ν_l^*, $\forall l \in E$ such that:

1. $\mathbf{x}^*$ and $\mathbf{y}^*$ are feasible,

2. $\nabla_{\mathbf{z}} L(\mathbf{z}^*, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)^T (\mathbf{z} - \mathbf{z}^*) \geq 0, \quad \forall \mathbf{z} \in \{\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}, \mathbf{x} \geq 0\}$.

The Lagrangian function results:

$$L(z, \lambda, \nu) = \sum_{l \in E} y_l D(y_l) + \sum_{s \in S} \lambda_s \left(f_s - \sum_{r:s(r)=s} x_r \right) + \sum_{l \in E} \nu_l \left(\sum_{r:l \in r} x_r - y_l \right) \tag{3}$$

When we differentiate it, we obtain:

$$\frac{\partial L}{\partial y_{\bar{l}}} = D(y_{\bar{l}}) + y_{\bar{l}} D'(y_{\bar{l}}) - \nu_{\bar{l}} \qquad\qquad \frac{\partial L}{\partial x_{\bar{r}}} = -\lambda_{s(\bar{r})} + \sum_{l \in \bar{r}} \nu_l \tag{4}$$

Observe that we can both increase and decrease $y_{\bar{l}}^*$ without going out of the polyhedral set X, i.e both $\begin{pmatrix} \mathbf{x}^* \\ \mathbf{y}^* \end{pmatrix} + \Delta y_{\bar{l}} \begin{pmatrix} \mathbf{0} \\ \mathbf{e}_{\bar{l}} \end{pmatrix}$ and $\begin{pmatrix} \mathbf{x}^* \\ \mathbf{y}^* \end{pmatrix} - \Delta y_{\bar{l}} \begin{pmatrix} \mathbf{0} \\ \mathbf{e}_{\bar{l}} \end{pmatrix}$ belong to X, where $\mathbf{e}_{\bar{l}}$ is the vector with the $\bar{l}$-th component equal to one and all the others equal to 0. Then, from the condition on the gradient of the Lagrangian, it follows:

$$\left. \frac{\partial L}{\partial y_{\bar{l}}} \right|_{\mathbf{z}^*} \Delta y_l \geq 0,$$

$$\left. \frac{\partial L}{\partial y_{\bar{l}}} \right|_{\mathbf{z}^*} \times (-\Delta y_l) \geq 0.$$

From these two results we can see that $\frac{\partial L}{\partial y_l}$ is equal to zero.

If $x_{\bar{r}}^* > 0$ it is possible to both increase and decrease it. In such case we get as above:

$$\left. \frac{\partial L}{\partial x_{\bar{r}}} \right|_{\mathbf{z}^*} = 0.$$

Instead if $x_r^* = 0$, it is only possible to increase it and we get:

$$\left. \frac{\partial L}{\partial x_{\bar{r}}} \right|_{\mathbf{z}^*} \geq 0$$

From (4), these conditions can be rewritten as follows in terms of the Lagrange multipliers:

$$\lambda_s^*(r) \begin{cases} = \sum_{l \in r} \nu_l^* & \text{if } x_r > 0, \\ \leq \sum_{l \in r} \nu_l^* & \text{if } x_r = 0, \end{cases} \tag{5}$$

$$\nu_l^* = D(y_l) + y_l D'(y_l). \tag{6}$$

We can interpret these equations as follows. First, each multiplier ν_l^* can be interpreted as a cost associated to link l, that is a function of the total traffic on the link. We can then associate a cost to each route as the sum of the link costs. Equation (5) then tells us that for each source-destination pair s all the routes used have the same route cost equal to λ_s^* and this is the minimum route cost (all the routes that are not used have higher route cost).

The optimal solution can then be achieved in a distributed way similarly to our initial naive solution. Costs for each link can be computed, aggregated over each route and propagated to each agent, who should then choose the route with minimum cost. The difference is in the right cost to consider: the link cost is composed of the delay (i.e. $D(y_l)$) plus the additional term $y_l D'(y_l)$.
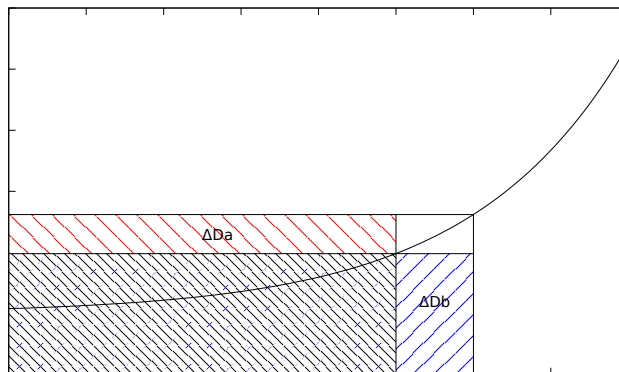

Figure 5: Additional term representation

In order to interpret the meaning of this additional term, let us consider what is the effect of adding some traffic Δy to a given link l. In Figure 5 the initial link delay experienced by all the users is denoted by the black area. When computing the new total delay, we need to consider the red area, ΔD_a and the blue area, ΔD_b, where $\Delta D_a = y_l D'(y_l)\Delta y$ and $\Delta D_b = \Delta y D(y_l)$. We observe that ΔD_b is the total delay experienced by this additional traffic and this is what the corresponding users would take into account in their choice to move to link l. Instead, ΔD_a is the total delay increase for the traffic already on link l. Then, considering the additional term $y_l D'(y_l)$ in the link cost corresponds to each user taking into account the delay increase that his choice would cause to the other users.

While we can assume that each user is immediately sensitive to the delay he experiences, we can wonder how we can make him responsive also to the additional term $y_l D'(y_l)$. A possibility in a transportation scenario is to add some tolls, so that drivers are encouraged to have a more desirable behaviour. We can revisit the example in the previous section and add the toll $y_l D'(y_l)$ at the cost of each link. Each user experiences then two new costs for traveling the two paths. $C_{upper} = 1 + y_{upper}D'(y_{upper}) = 1$ and $C_{lower} = y_{lower} + y_{lower}D'(y_l) = 2y_{lower}$. Now the drivers choose the link where the total cost, delay plus toll, is minimum and the final solution is that half of the drivers choose the lower link and the others choose the upper link, that is the optimal solution for Problem (1).

References

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