Applied Nonlinear Optimization

SF2822

Project Assignment 2: 2A

Group 2A1

Isak Maxstad
Daniel Liu
San Khaffaf

Stockholm

11 May 2021

Abstract

In this report, we formulate a model for optimizing the total time it takes for the population of Stockholm to travel to work. Two cases are considered. First, we consider the case where commuters choose the fastest path without regard to congestion. Then, we consider the case where commuters consider congestion when choosing their route. The results show that if every commuter chooses the fastest path, the total travel time would be 15.7 million minutes. In comparison, if every commuter optimized their route with regard to congestion, the total time would be 14.3 million minutes. The population of Stockholm would save almost 1.5 million minutes by considering congestion when planning their route. Further, we also present a mathematical basis for choosing which roads to improve. Lastly, we discuss the limitations and the potential improvements of the model.

1 Background

Every day, the people of Stockholm need to drive to work. When planning the route, one obviously wants to take the shortest route. However, the shortest route is not always the fastest route, since congestion can slow down popular roads considerably. The purpose of this paper is to find the fastest route between any two regions around the Stockholm area. The objective is to minimize the sum of the traveling time for every person in Stockholm. First, we find the fastest route without including congestion for every person in Stockholm regardless of their origin and destination. After the optimal routes are found, we further include congestion in the calculations and find the fastest route with respect to congestion. These two routes are compared with each other, and the change of traffic flow is visualized. We also aim to give a recommendation of which roads to expand, and quantify the improvement of the change. The models are formulated as nonlinear optimization problems using GAMS.

1.1 Description of Regions

We assume that Stockholm has seven regions: Sollentuna, Täby, Solna, Centrum, Södermalm, Nacka, and Huddinge. Four of these regions have roads directly to Centrum, while two of these do not have roads directly connected to Centrum. A map of the regions of Stockholm and the connected roads are shown in Figure 1. To travel between two regions that are not directly connected, one has to drive through another region, and then use the roads that emancipate from that region.

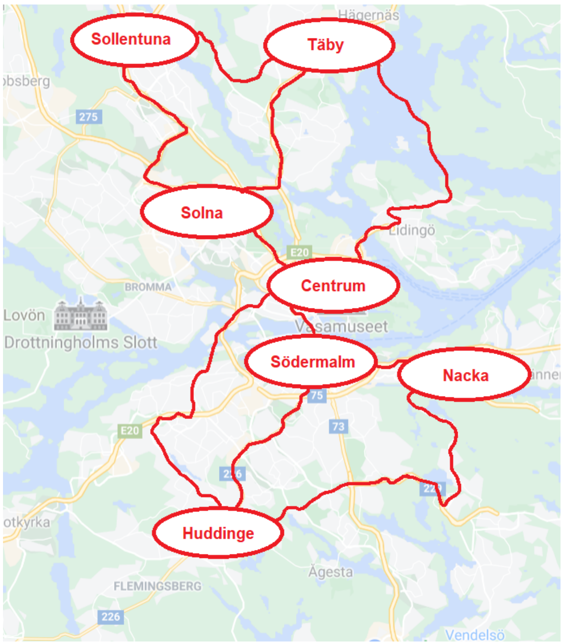

Figure 1: *Map of Stockholm and its suburbs*

There are people who live in every region. There are also businesses that operate in every region, and we assume that the workers in each region consist of people from every other region. It can be difficult to precisely map how many people who live in one region work in another region. For the purpose of this paper, we assume that the share of a regions' population that work in another region is proportional to the percentage of GDP of the other region. The population and the percentages of GDP are shown in Table 1.

Table 1: *The population and GDP share of each region*

Region	Population (thousands)	Percentage of GDP
Solna	82	10
Nacka	105	5
Sollentuna	74	5
Täby	65	5
Södermalm	130	20
Huddinge	113	5
Centrum	347	50

As an example, Solna has a population of 82 000, while Nacka has a GDP share of 5 percent. This implies, by our model, that the number of people who work in Nacka and who live in Solna is $82\ 000 \cdot 5\% = 4100$. The distribution of the population for each region is shown in Table 2. Of course, there are some people who live and work in the same region. These people do not need to drive to work, thus they are excluded from the model. Their value in the table is set to 0.

Table 2: *The number of people moving between regions (thousands)*

Origin \ Destination	Solna	Nacka	Sollentuna	Täby	Södermalm	Huddinge	Centrum
Solna	0	4.1	4.1	4.1	16.4	4.1	41
Nacka	10.5	0	5.25	5.25	21	5.25	52.5
Sollentuna	7.4	3.7	0	3.7	14.8	3.7	37
Täby	6.5	3.75	3.75	0	13	3.75	37.5
Södermalm	13	6.5	6.5	6.5	0	6.5	65
Huddinge	11.3	5.65	5.65	5.65	22.6	0	56.5
Centrum	34.7	17.4	17.4	17.4	69.4	17.4	0

In this model, the only roads that exist are the ones marked in Figure 1. We assume that if a road exists in one direction, it also exists in the other direction. Furthermore, we assume that the the travel time when there is no traffic is given in Table 3. When assigning these values, we assume that the time is proportional to the distance between nodes.

Table 3: *Time between regions when there is no traffic (minutes)*

	Solna	Nacka	Sollentuna	Täby	Södermalm	Huddinge	Centrum
Solna			13	12			6
Nacka					15	23	
Sollentuna	13			14			
Täby	12		14				15
Södermalm		15				13	3
Huddinge		23			13		15
Centrum	6			15	3	15	

2 Mathematical Formulation

For the mathematical model, we consider the map of Stockholm as a graph where the nodes are the regions and the arcs are the roads between regions. The arcs are directed in both directions, since one is able to travel in any direction between nodes. Furthermore, when modelling with congestion, the congestion of a road in one direction is seen as separate from congestion in the other direction. Lastly, we view congestion as a continuous flow that always exists. We do not register the location of each car during every moment; If a car travels on a route, we assume that it contributes to congestion simultaneously during the whole day. In reality, congestion is more severe during rush hour than in the middle of the day.

2.1 Optimization Program without Regard to Congestion

In the simplest model, we choose the fastest route for every person in Stockholm, without considering congestion. We want to make sure that every commuter arrives precisely at their desired destination, but we also want to know which route every person takes. For this, we need to classify the commuters on each road into groups depending on their origin and destination. We define a set of variables $x_{(i,j,k,l)}$ representing the number of cars (we equate the number of people with the number of cars), where

- i: At which node the people live,
- j: At which node the people work,
- k: At which node the arc starts,
- l: At which node the arc ends.

In other words, $x_{(i,j,k,l)}$ is to be read as the number of cars that want to go from node i to node j that are on the arc from node k to node l. The indices $i, j, k,$ and l are all representations of the same set (the nodes). It is rather the order in which they are written that matters. For clarity, we in this report always denote an index as i for the origin, j for the destination, k for where the road starts, and l for where the road ends. If we want to represent an arbitrary node, we refer to it with index v.

One possible objective function is to minimize the sum of all travel times. Let $t_{(k,l)}$ denote the time to travel on an arc between k and l without considering congestion. The value of these parameters are given in Table 3. The sum of all travel times is then given by

$$\sum_{i,j,k,l} t_{(k,l)} x_{(i,j,k,l)}.$$

For the constraints, we partition the whole feasible set into several subsets. We model the constraints for the number of cars separately for each pair of origin and destination nodes (i, j). Given an origin node and a destination node, the problem of finding the fastest route between two nodes can be modeled as a network problem. For a network problem, the sum of all incoming arcs to a node should be equal to all of the outgoing arcs of the node. This way, the system is in balance and no cars are lost from the system. Two exceptions are when cars leave from their origin node, and when they arrive at their destination node. When cars leave from their origin node, they are added to the system and considered as inputs to the system. Likewise, when cars arrive at their destination node, they are removed from the system, considered as outputs. This relationship can be written as one single equation:

$$\sum_{k} x_{(i,j,k,v)} + p_{(i,j,v)} = \sum_{l} x_{(i,j,v,l)} + q_{(i,j,v)}, \ \forall i, j, v.$$

For each selected node v, $p_{(i,j,v)}$ is equal to the number of commuters who want to travel from node i to node j if the selected node is the origin node, and 0 otherwise. Likewise $q_{(i,j,v)}$ is equal to the number of commuters from node i to node j if the selected node is the destination node, and 0 otherwise. $p_{(i,j,v)}$ and $q_{(i,j,v)}$ can be written as

$$p_{(i,j,v)} = \begin{cases} \text{numbers of commuters from node } i \text{ to } j, & \text{if } i = v \\ 0, & \text{otherwise} \end{cases}$$

and

$$q_{(i,j,v)} = \begin{cases} \text{numbers of commuters from node } i \text{ to } j, & \text{if } j = v \\ 0, & \text{otherwise} \end{cases}$$

The approach of using only one equation as constraint solves three problems. First, for nodes that are neither the origin nor the destination, the input to the node is equal to the output from the node. Second, for nodes that are origin or destination nodes, the number of cars that either leaves from their origin or arrives at their destination are accounted for when maintaining the balance of the system. Third, it prevents the optimization program from choosing a route that loops back to the origin, since arriving at the origin node disrupts the balance of that node, forcing the route to leave the node again. This should never be never optimal as long as the costs of traveling are positive.

Lastly, we require the number of cars to be greater than or equal to 0. In this problem, we assume that $x_{(i,j,k,l)}$ is sufficiently large to be considered a continuous variable. The full optimization program is formulated as

$$\begin{aligned} &\text{minimize} && \sum_{i,j,k,l} t_{(k,l)} x_{(i,j,k,l)} && \\ &\text{subject to} && \sum_{k} x_{(i,j,k,v)} + p_{(i,j,v)} = \sum_{l} x_{(i,j,v,l)} + q_{(i,j,v)}, && \forall i,j,v, \\ & && x_{(i,j,k,l)} \geq 0, && \forall i,j,k,l. \end{aligned} \tag{1}$$

The solution of Equation (1) should give separate shortest-path suggestions for each commuter. This can be verified by decomposing the problem into subproblems for each origin node i and destination node j separately, and noting that the solution is the same as the solution from Equation (1). The optimal value of the problem should be equal to $\sum_{i,j} z_{(i,j)}$, where

$$\begin{aligned} z_{(i,j)} = \quad &\text{minimize} && \sum_{k,l} t_{(k,l)} x_{(k,l)}, && \\ &\text{subject to} && \sum_{k} x_{(k,v)} + p_{(i,j,v)} = \sum_{l} x_{(v,l)} + q_{(i,j,v)}, && \forall v, \\ & && x_{(k,l)} \geq 0, && \forall k,l. \end{aligned} \tag{2}$$

One advantage with the formulation in Equation (1) is that it saves all of the optimal routes for all of the origins and destinations. With the solution, we can trace precisely the optimal route for every origin and destination. However, the optimization program is also very large and becomes more difficult to compute as the number of nodes increases. The number of variables for this problem is of order $\mathcal{O}(n^4)$, where n denotes the number of nodes. In comparison, the number of variables for the decomposed problem is only $\mathcal{O}(n^2)$, which is much easier to solve. However, we would then only know the optimal value and perhaps the total number of cars on each road, but not precisely the origin and destination of each commuter.

2.2 Optimization Program with Regard to Congestion

In order to consider more realistic conditions, we need to consider the traffic congestion on the roads. This case examines the effects congestion will have on the total time it takes for all cars to arrive at their destinations. In contrast to the previous program, the result does not necessarily return the shortest path for each individual as the optimal solution. Due to congestion, certain individuals traveling from and to the same locations may take different paths in order to reduce the number of cars on each arc.

To account for congestion, the time associated with traveling on each arc must be modified. As a result of this, the objective function is modified as presented in Equation (3). The objective function introduces two new positive parameters $u_{(k,l)}$ and $a_{(k,l)}$ as well as one new variable $y_{(k,l)}$. As seen in the constraints of the program, variable $y_{(k,l)}$ denotes the total number of cars that use arc (k,l) regardless of their origin and final destination. This value is restricted by the maximum capacity of cars on each road, $u_{(k,l)}$. The constant $a_{(k,l)}$ scales the increase in time due to congestion. Different roads filled to a certain percentage of their capacity should not see the same valued increase in time as a result of differences in traffic lights, roundabouts etc., this is represented in $a_{(k,l)}$.

$$\begin{aligned}
\text{minimize} \quad & \sum_{k,l}(t_{(k,l)} + \frac{a_{(k,l)}y_{(k,l)}}{u_{(k,l)} - y_{(k,l)}})y_{(k,l)} & \\
\text{subject to} \quad & \sum_{k} x_{(i,j,k,v)} + p_{(i,j,v)} = \sum_{l} x_{(i,j,v,l)} + q_{(i,j,v)}, & \forall i,j,v, \\
& y_{(k,l)} \leq u_{(k,l)}, & \forall k,l, \\
& y_{(k,l)} = \sum_{i,j} x_{(i,j,k,l)}, & \forall k,l, \\
& x_{(i,j,k,l)} \geq 0 & \forall i,j,k,l.
\end{aligned} \tag{3}$$

As there is no exact value of the capacity of each road available to us, the determined values of capacity on each arc, as presented in Table 4, are mere estimations. The determining factors when choosing the values of $u_{(k,l)}$ included the size of the populations of regions connected by the arcs, the number of cars traveling on the roads in the previous program and symmetry of each arc in both directions was assumed. Realistically, arcs connecting larger regions in a city would be constructed to have a higher capacity of cars, thus, this is reflected in the chosen values of capacity. We want traffic congestion to increase the travel time and result in some cars choosing different paths. However, limiting the capacities to reflect values close to the number of commuters on all arcs in the previous program would result in the time increasing towards infinity, since the roads would be full. Therefore, the capacity of each arc was chosen to be at least twice the amount of commuters in the program without congestion in all cases apart from between Södermalm and Centrum.

Table 4: *Maximum capacity of cars on each arc, $u_{(k,l)}$ (thousands)*

	Solna	Nacka	Sollentuna	Täby	Södermalm	Huddinge	Centrum
Solna			180	100			270
Nacka					200	200	
Sollentuna	180			160			
Täby	100		160				170
Södermalm		200				160	220
Huddinge		200			160		270
Centrum	270			170	220	270	

To determine values of the positive constant $a_{(k,l)}$, the equation for the time on each arc was considered for cases where the arcs were filled to half their capacities. In these cases the travel time on the arc increased by the value of $a_{(k,l)}$ due to congestion. With regards to the values of the original travel times on the arcs, an increase of 4 minutes seemed to be a reasonable estimate for most roads and was thus chosen. However, in cases where the roads have several intersections, roundabouts or stoplights it was assumed that despite the same percentage of capacity filled, these would further increase the travel time. Thus, an increased value of $a_{(k,l)}$ was assumed for certain arcs, mainly arcs passing through Södermalm and Centrum due to the large number of intersections here. The chosen values for the set of parameters, $a_{(k,l)}$, are displayed in Table 5.

Table 5: *Values of constant $a_{(k,l)}$ (minutes)*

	Solna	Nacka	Sollentuna	Täby	Södermalm	Huddinge	Centrum
Solna			4	4			5
Nacka					5	4	
Sollentuna	4			4			
Täby	4		4				4
Södermalm		5				4	6
Huddinge		4			4		4
Centrum	5			4	6	4	

2.3 Global Optimality of Optimization Programs

In order to determine whether the optimal solutions obtained in the two optimization programs, one can consider whether the two programs are convex. If the programs are convex, the local minimizers obtained are also global minimizers. Investigating convexity of the programs, it is sufficient to determine if the feasible regions of the problems constitute a convex set and if the objective function is convex on the feasible region.

In the program without regard for congestion, the two constraints constituting the feasible region form a convex set. This can be observed as the both constraints in Equation (1) are linear. The objective function is a sum of linear equations on the convex set, thus, the optimization program is convex and as a result, the optimal solution is the global minimizer.

The optimization program that takes congestion into account has two constraints that remain from the previous program. As seen in Equation (3), the first and last constraint remain and as stated, these do not hinder convexity of the set. The two additional constraints in the program are both linear, and thus the feasible region in the program considering congestion is also a convex set. To investigate whether the modified objective function is convex within this region, the second derivative with respect to y can be evaluated. If the function has a positive second derivative on the region, it is convex. The terms of the objective function and its derivatives are presented in Equation (4).

$$
\begin{aligned}
f(y_{(k,l)}) &= (t_{(k,l)} + \frac{a_{(k,l)}y_{(k,l)}}{u_{(k,l)} - y_{(k,l)}})y_{(k,l)} \\
f'(y_{(k,l)}) &= t_{(k,l)} + \frac{a_{(k,l)}y_{(k,l)}(2u_{(k,l)} - y_{(k,l)})}{(u_{(k,l)} - y_{(k,l)})^2} \\
f''(y_{(k,l)}) &= \frac{2a_{(k,l)}y_{(k,l)}^2}{(u_{(k,l)} - y_{(k,l)})^3} + \frac{4a_{(k,l)}y_{(k,l)}}{(u_{(k,l)} - y_{(k,l)})^2} + \frac{2a_{(k,l)}}{u_{(k,l)} - y_{(k,l)}}
\end{aligned}
\tag{4}
$$

The second derivative of the objective function is positive. This holds as the constants $a_{(k,l)}$ and $u_{(k,l)}$ are given to be positive for all arcs, and the variable $y_{(k,l)}$ is both constrained to be smaller than $u_{(k,l)}$ and positive as it is a sum of the positive variable $x_{(i,j,k,l)}$ over i and j. Thus, the objective function is convex on a convex set, the optimization program is convex and the optimal solution obtained is a global minimizer.

2.4 A Mathematical Basis for Improving Roads

When choosing which road to improve, one needs to consider the economic and the technical limitations of the desired improvement. For example, there may be a fixed budget, and the improvement of the roads are associated with a cost, or if there are physical or legal obstacles that prevent the expansion of a particular road. Without this kind of information, we can only make recommendations for hypothetical scenarios.

We can calculate the marginal improvement of increasing the capacity of each $u_{(k,l)}$ at the point of the optimal solution in Equation (3). If we take the partial derivative of the objective function with respect to each $u_{(k,l)}$, we get the expression

$$\frac{\partial}{\partial u_{(k,l)}} \sum_{k,l} (t_{(k,l)} + \frac{a_{(k,l)}\bar{y}_{(k,l)}}{u_{(k,l)} - \bar{y}_{(k,l)}})\bar{y}_{(k,l)} = -\frac{a_{(k,l)}\bar{y}^2_{(k,l)}}{(u_{(k,l)} - \bar{y}_{(k,l)})^2} \tag{5}$$

where $a_{(k,l)}$ is a constant, $u_{(k,l)}$ is the capacity of the road, and $\bar{y}_{(k,l)}$ is the (fixed) solution of Equation (3). If we investigate the function we see that the biggest impact the capacity has on the travel time is when the number of cars traveling on the road is approaching the road capacity. The derivative will at that point move towards negative infinity. Since the denominator is quadratic, the increase in capacity will impact the travel time the most when the road is almost near full capacity and become less impactful quickly as the roads become less crowded.

Thus, one way to choose how to best improve capacity of the system is to choose the road which has the largest negative derivative in our current optimal solution. It is difficult to say how much higher the capacity needs to be, but one way is to increase the capacity of the road with the most negative derivative until the marginal gain of further improving the road is less than the next best road to improve. It is also difficult to estimate the effect on the optimal value, since the value of the derivatives are only valid approximations of the effect of the improvement for points near the optimal solution of Equation (3).

To get a more accurate estimation, we can expand the objective function into Taylor series of higher order. Let $h(u)$ denote one term of the objective function at the optimal value. If we suppose we want to improve the capacity of one road by Δu. Then, the second order Taylor polynomial at $h(u + \Delta u)$ is

$$h(u + \Delta u) \approx h(u) + h'(u)\Delta u + \frac{h'(u)}{2}(\Delta u)^2,$$

where $h'(u)$ and $h''(u)$ are the first and second derivatives of $h(u)$ respectively, with respect to u. Thus, an approximated improvement of the objective function is

$$h(u + \Delta u) - h(u) \approx h'(u)\Delta u + \frac{h'(u)}{2}(\Delta u)^2.$$

Another approach of choosing which roads to improve can be used when the total capacity of the roads can only be increased by a fixed amount. In this scenario, we assume that there is a possibility to distribute the improvements arbitrarily, and that the potential cost of the improvements would be proportional to the capacity increase. We can then set up an optimization program which solves for which roads and by how much they should be improved. If we suppose that we want to minimize the sum of all traveling time with respect to congestion, then this problem can be modeled with some minor changes from Equation (3).

Let $c_{(k,l)}$ be a variable that denotes the capacity improvement for the road between node k and node l, and c_{max} be a constant that denotes the total amount of capacity increase we have budgeted for. Then, the optimization program can be formulated as

$$\begin{array}{lll}
\text{minimize} & \sum_{k,l} (t_{(k,l)} + \frac{a_{(k,l)} y_{(k,l)}}{u_{(k,l)} + c_{(k,l)} - y_{(k,l)}}) y_{(k,l)} & \\
\text{subject to} & \sum_{k,l} c_{(k,l)} \leq c_{max} & \\
& \sum_{k} x_{(i,j,k,v)} + p_{(i,j,v)} = \sum_{l} x_{(i,j,v,l)} + q_{(i,j,v)}, & \forall i,j,v, \\
& y_{(k,l)} \leq u_{(k,l)}, & \forall k,l, \\
& y_{(k,l)} = \sum_{i,j} x_{(i,j,k,l)}, & \forall k,l, \\
& x_{(i,j,k,l)} \geq 0, & \forall i,j,k,l, \\
& c_{(k,l)} \geq 0, & \forall k,l.
\end{array} \tag{6}$$

The only difference between Equation (3) and Equation (6) is that the expression for the fixed capacity $u_{(k,l)}$ in Equation (3) is replaced with an expression for the fixed capacity plus potential improvements in Equation (6), which need to be positive (we cannot reduce the capacity of roads) and less than or equal to the total road improvement budget.

Of course, it it usually not possible to distribute the improvement arbitrarily between roads. A more practical approach may be to improve a discrete number of roads with a fixed capacity. Let c be the fixed capacity increase of one road improvement, and $w_{(k,l)}$ be a variable that decides the number of improvements made on the road from k to l. If m is the number of road improvement the budget allows for, then the we can formulate an optimization program as

$$\begin{array}{lll}
\text{minimize} & \sum_{k,l} (t_{(k,l)} + \frac{a_{(k,l)} y_{(k,l)}}{u_{(k,l)} + c w_{(k,l)} - y_{(k,l)}}) y_{(k,l)} & \\
\text{subject to} & \sum_{k,l} w_{(k,l)} \leq m & \\
& \sum_{k} x_{(i,j,k,v)} + p_{(i,j,v)} = \sum_{l} x_{(i,j,v,l)} + q_{(i,j,v)}, & \forall i,j,v, \\
& y_{(k,l)} \leq u_{(k,l)}, & \forall k,l, \\
& y_{(k,l)} = \sum_{i,j} x_{(i,j,k,l)}, & \forall k,l, \\
& x_{(i,j,k,l)} \geq 0 & \forall i,j,k,l, \\
& w_{(k,l)} \geq 0 \text{ and integer}, & \forall k,l
\end{array} \tag{7}$$

.

This approach can be modified and specified even more as more information of the preconditions are known. For example, improving different roads may be associated with different costs. The constant c could then be replaced with a parameter that is different for different roads. There is also an option of improving the parameter $a_{(k,l)}$ instead of $u_{(k,l)}$, which could include optimizing traffic lights and speed limits. In any case, the problem is formulated analogously by adding a variable that represents the change in the parameter.

3 Results and Analysis

3.1 Visualization of Traffic Flows

Solving the two optimization programs in Equation (1) and (3) provides the results as presented in Table 6. In order to clearly see the differences in paths taken for individuals in the regions one can observe Figure 2, where the thickness of the arrows is proportional to the number of vehicles using each arc. The minimal total time for the whole population is approximately 9.40 million minutes on the roads when all drivers take the shortest path without considering congestion and 14.3 million minutes with congestion taken into account. The difference in total time is significant However, this was to be expected as the inclusion of congestion will increase the time per car on arcs taken regardless of whether the cars take a different path or not. Instead, what is more important to note is the differences in paths taken and the time gained for the population as a whole by certain cars choosing to take longer paths to their destinations. If all cars had maintained the shortest paths they took in the original program but with congestion considered, as would be most realistic, the total time would be approximately 15.7 million minutes. Thus, almost 1.5 million minutes are gained among the population if certain people choose to take longer paths to their destinations.

Table 6: *Number of cars using arc (k, l) (thousands) and total time (thousand minutes) for all cars on the arc for original program and program accounting for congestion*

Arc	**Without Congestion**		**With Congestion**	
	$y_{(k,l)}$	Time, T	$y_{c(k,l)}$	Time, T_c
Centrum - Huddinge	34.9	524	50.8	810
Centrum - Solna	104	626	104	953
Centrum - Södermalm	142	426	120	1210
Centrum - Täby	34.8	521	34.8	557
Huddinge - Centrum	79.1	1190	118	2140
Huddinge - Nacka	5.65	130	28.1	664
Huddinge - Södermalm	22.6	294	22.6	309
Nacka - Huddinge	5.25	121	44.1	1070
Nacka - Södermalm	94.5	1420	55.6	942
Sollentuna - Solna	66.6	866	60.1	902
Sollentuna - Täby	3.70	51.8	10.2	145
Solna - Centrum	125	749	118	1180
Solna - Sollentuna	38.9	505	38.9	548
Solna - Täby	4.10	49.2	4.10	49.9
Södermalm - Centrum	171	513	126	1380
Södermalm - Huddinge	0	0	6.50	85.6
Södermalm - Nacka	34.9	524	12.5	191
Täby - Centrum	52.0	780	58.5	1000
Täby - Sollentuna	6.50	78.0	3.25	45.8
Täby - Solna	3.25	45.5	6.5	79.8
Total		9405		14250

The most significant differences in behavior of the cars with congestion taken into account occurs in the southern regions. Observing Figure 2, one can see distinct differences in the thickness of arrows in these regions. It is more beneficial for cars to travel on the arc between Södermalm and Centrum to a lesser extent. The original problem resulted in a large number of cars traveling on this arc, however, as the capacity of the arc was close to this value and the constant $a_{(k,l)}$ was high, the congestion would cause much longer travel times. Instead, the optimal solution would be for more cars to travel through the arc between Huddinge and Centrum when going south or north. The congestion leads to it being more beneficial for a group of cars to travel from Södermalm to Huddinge, an arc that was not used in this direction when congestion was not taken into account. The three northernmost regions see small variations in number of cars on each arc. This can be explained by the high capacity of arcs compared to the low number of commuters through these regions as well as the similarity in populations, travel times and number of commuters between the three.

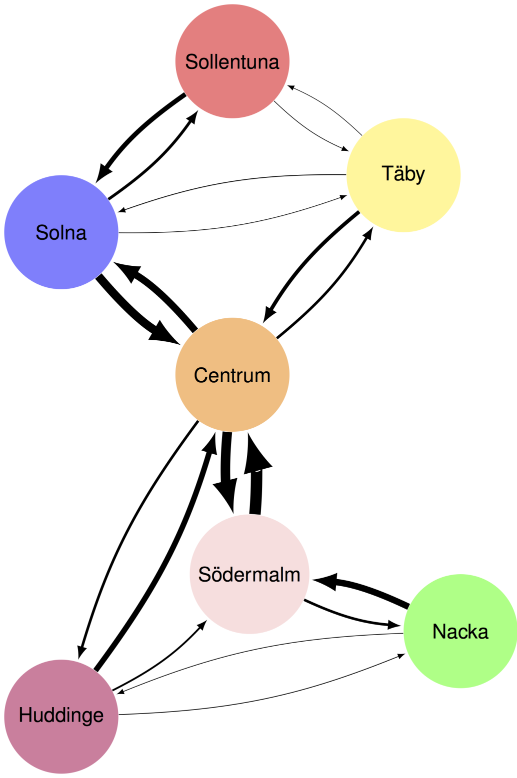

Figure 2: *Graph displaying paths cars choose in the original program (left) and the program taking congestion into account (right). The thickness of the arrows signifies the number of cars traveling on the arc*

3.2 The Optimal Road to Improve

As mentioned in Section 2.4, the amount and the number of roads which are able to be improved is determined by economic and technical constraints. In this problem, we suppose that we want to improve one road with a capacity of 50 thousand.

In the solution with regard to congestion we see in Figure 2 that the road between Södermalm and Centrum has the highest traffic, which hints that it is beneficial to increase the capacity there. A better guideline is given in Table 7, where the first and second derivative with respect to $u_{(k,l)}$, denoted h' and h'' respectively, for a term in the objective function are shown, along with the estimated increase a second order Taylor expansion near the objective value $\bar{y}_{(k,l)}$. The table shows that the road between Södermalm and Centrum has the largest estimated reduction to the objective function near the optimal solution.

Table 7: *The roads which has the largest negative marginal reduction to the objective value if improved, near the optimal solution*

Arc	$\bar{y}_{(k,l)}$	$a_{(k,l)}$	$u_{(k,l)}$	h'	h''	$50 \cdot h' + 250 \cdot \frac{h''}{2}$
Södermalm - Centrum	126	6	220	-10.63	0.23	-503.4
Centrum - Södermalm	120	6	220	-8.51	0.17	-404.2
Solna - Centrum	118	5	270	-3.04	0.04	-147.1
Huddinge - Centrum	118	4	270	-2.41	0.03	-116.5
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

We can test our hypothesis by manually increasing the capacity for a certain road in the optimization program, and comparing the optimal value with our predicted value. Table 8 shows the predicted improvement compared to the actual improvement for some roads. The table shows that the predicted improvement is always too large. We discuss this in Section 2.4, where we argue that a improvement is most effective when the road is near full capacity, and thereafter the marginal reduction of the objective value becomes less effective as the road becomes less full.

Table 8: *The predicted and actual reduction in objective value for the improvement of some roads (thousand minutes)*

Arc	Predicted Improvement	Actual Improvement
Södermalm - Centrum	-503.4	-443.4
Centrum - Södermalm	-404.2	-379.1
Solna - Centrum	-151.0	-147.1
Huddinge - Centrum	-116.5	-94.0

Even if the predicted reduction of the optimal value is not very accurate for a road improvement of 50 thousand cars, the ranking of the potential roads to improve is not changed from the predicted case. Thus, our recommendation is to increase the road between Södermalm and Centrum. This is further confirmed by solving the optimization program formulated in Equation (7), which also recommends improving the road from Södermalm to Centrum.

4 Discussion

The models presented in this report are simplifications of reality. For example, we have assumed that every person owns one car and thus only looked at the travel time by car. Other interesting aspects would be to include buses and trains as well. One could assume that it often takes a bit longer to travel by bus with all the stops, but if everyone traveled by bus there would be less congestion. Other interesting aspects when including buses and trains is to look at the cost for traveling with toll fees and fuel or looking at emissions and pollution.

A mathematical formulation with consideration to buses could be formulated as in Equation (8) where the outlines are similar to the ones in Equation (3). The new parameters are y_{pe} which are the total number of people traveling in Stockholm, y_{ve} which are the total number of vehicles on the roads and b which is the number of people taking the bus. There are also two more parameters: C_b, which is the number seats in one bus and t_b, which is the extra time it takes for a bus to travel compared to a car when there is no congestion. Two more constraints are also added, where we assume that the number of people traveling by bus is between 20-40% of the total population.

$$
\begin{aligned}
\text{minimize} \quad & \sum_{k,l}((t_{(k,l)} + \frac{a_{(k,l)} y_{ve(k,l)}}{u_{(k,l)} - y_{ve(k,l)}}) y_{pe(k,l)} + t_{b(k,l)} \sum_{i,j} b_{(i,j,v,l)}) & & \\
\text{subject to} \quad & \sum_{k} x_{(i,j,k,v)} + b_{(i,j,k,v)} + p_{(i,j,v)} = & & \\
& \sum_{l} x_{(i,j,v,l)} + b_{(i,j,v,l)} + q_{(i,j,v)}, & & \forall i, j, v, \\
& y_{ve(k,l)} \leq u_{(k,l)}, & & \forall k, l, \\
& y_{ve(k,l)} = \sum_{i,j}(x_{(i,j,k,l)} + \frac{b_{(i,j,k,l)}}{C_b}), & & \forall k, l, \\
& y_{pe(k,l)} = \sum_{i,j}(x_{(i,j,k,l)} + b_{(i,j,k,l)}), & & \forall k, l, \\
& x_{(i,j,k,l)} \geq 0, & & \forall i, j, k, l, \\
& b_{(k,l)} \geq 0.2 \cdot y_{pe(k,l)}, & & \forall k, l, \\
& b_{(k,l)} \leq 0.4 \cdot y_{pe(k,l)}, & & \forall k, l.
\end{aligned} \tag{8}
$$

Another simplification made in the models is that we assume that everyone is on all of the roads simultaneously. In reality, certain roads are likely more crowded early in the morning when everyone leaves home, while other roads are more crowded when people are about to arrive at work. To model a more accurate representation of reality, the time period should be taken into consideration.

Adding more variables makes the model a better representation of reality, but as more variables are added, the computation capacity becomes a problem. The model is already large and grows quickly with every node added. In our model, the only way to travel from the northern side to the southern side is to drive through Centrum. In this case, another solution would be to look at the map as two separate halves since everyone from the north of Centrum needs to go through Centrum and vice versa for people from the south of Centrum. The south side would be solved as in our current model with the simplification that everyone that would travel to a suburb in the north is now considered to be commuting to Centrum. In the second step we optimize the best way from Centrum to the northern suburbs. In parallel with this, the same thing is done but for the northern part of the city. In total, four optimization problems need solved, but by halving the number of nodes, each problem only has around $\frac{1}{16}$ of the variables as in the original problem. Our expectation is that this method is less expensive to compute.

Another way to make the computation cheaper could be to solve each pair of start and end nodes iteratively instead of solving them all at once. For example, we could calculate the best solution for all commuters between one pair of nodes and then use that solution as a condition for the next iteration. The model would then find an optimal route for the next pair of nodes given the cars already assigned to roads. This would solve many small problems, which often is less computationally intensive than solving one large problem. However, with this method the first pair to be assigned would have the best solution and the solution for every following region would be a little bit worse than the previous one, which could be seen as unfair for the commuters.

However unfair, the approach of letting some commuters have priority on roads may not be too far away from reality. If a commuter departs to work earlier than everyone else, then they would experience less congestion on the roads. Using for example GPS data, we can know how many cars are on the road and find an optimal route for a commuter who is about to leave home, regardless of when they want to leave. In other words, we can solve an optimization problem for every individual which depends on the time they leave for work and the status of the other cars on the roads. Computationally, this is on the other extreme of solving one very large problem, which may not necessarily be computationally efficient. However, the time period in which these problems need to be solved spans over a whole day, which means that it can still be useful.

5 Conclusion

Analyzing the results provides an understanding of the importance and validity of the two optimization programs formulated, as well as the ability to draw conclusions regarding their applicability in real life. The original optimization program does not provide means to analyze the results and their applicability on its own. However, it does provide confidence in the mathematical models as it returns the shortest path for all commuters, and comparing the differences in paths taken between the two programs allows one to further understand where congestion would occur most, and more importantly which roads would be most beneficial to modify. If one were to analyse the behavior of real commuters, one could assume that most people would take the shortest path to their destination regardless of congestion on those roads.

In order to draw valid conclusions from the results one would need more information on the actual amount of travelers and more accurate values of road capacities and increased times based on the capacities. However, as the models are formulated, they allow for modifications in input values. This would increase the accuracy and, together with an inclusion of bus travelers and accounting for departure times, allow the same programs to be used for decision-making as they provide a sufficient overview of patterns of behavior.