Senior Thesis in Mathematics

The Parks Location Problem

Author:
Alan Antonio Peral Ortiz

Advisor:
Dr. Shahriar Shahriari

Submitted to Pomona College in Partial Fulfillment
of the Degree of Bachelor of Arts

April 8, 2018

Abstract

We take a look at the multiobjective facility location problem of park placement in the city of Bogotá, Colombia. We begin with an introduction to operations research, delving into its development and history, before we move on a to simplified example model. We work up from this example to look at the final model. We then examine more information about multiobjective and facility location problems.

Contents

Chapter 1

Introduction

1.1 What is Operations Research?

We begin with a brief history of Operations Research (OR), which as a modern discipline has its roots in the Soviet political economy as well as the British war effort of World War II.

[SECTION ABOUT SOVIET UNION + OR]

Now, in the year 1934 as Nazi Germany denounced the Treaty of Versailles, Britain sensed the threat and raced to strengthen its defenses. By the following year, radar had been developed and was capable of effectively detecting enemy aircraft, although its utility was stringent on its ability for integration with the existing defense systems: ground observers, interceptor aircraft, and antiaircraft artillery positions. The first task of the newly formed operational research group was to use scientific, rigorous processes to develop a system for the incorporation of radar into existing infrastructure, which was far from the more well-defined mathematical processes that exist today. However, this integrated radar-based air defense system increased the probability of intercepting an enemy aircraft by a factor of ten [CITE HISTORY OF OR IN US MILITARY BOOK].

Soon, the usage of OR spread to investigate other problems, and in 1940 was even called upon to influence high-level strategic policy, when the French requested additional RAF fighter support. Churchill was inclined to acquiesce to the request, but an OR team showed that sending more RAF fighters would weaken Britain beyond recovery in the face of a German attempt to invade Britain [CITE HISTORY OF OR IN US MILITARY]. Churchill was convinced by their presentation, and so he did not send the additional aircraft, preserving the pilots and aircrafts for the Battle of Britain instead. This decision, along with the incorporation of radar, contributed significantly to Britain's victory in the Battle of Britain [CITE HISTORY OF OR IN US MILITARY]. Another significant accomplishment of OR analysts involved the work of deciding upon depth charge settings for bombs. Their work in this field led to an immediate improvement on aerial attacks of German submarines, with estimates of the increased efficiency ranging from 400 to 700 percent [CITE HISTORY OF OR IN US MILITARY]. Other work included reducing the number of artillery rounds required to down one German aircraft from twenty thousand in the summer of 1941 to merely four thousand the following year [CITE HISTORY OF OR IN US MILITARY]. From the British, this discipline spread to the Americans, where it was also used in the war effort. [TALK MORE ABOUT 'MURICA HERE!].

There was a specific need to allocate scarce resources to military operations in an effective manner, and so the British and US militaries had scientists perform *research on* (military) *operations*. In essence, the goal was to make the war machine more efficient, and they succeeded. They developed effective ways to use the new radar technology, as well as came up with better ways to manage convoys and conduct antisubmarine operations [CITE INTRO TO OR BOOK].

The success that OR saw in the war then encouraged interest in nonmilitary applications of the field. A cursory glance at a variety of introductory textbooks will reveal that there is a certain focus on private sector applications of the field [Should probably cite this claim]. These applications tend to be primarily concerned with profit maximization and other aspects of running a business. While these problems provide for some interesting mathematical formulations, the field's ability to

There is a subfield of the discipline called Community-Based Operations Research, which seeks to shift the focus of OR from profit maximization or cost-reduction to improving the quality of life within a community. One of the main advocates of the field and of this lens that focuses on people as opposed to money is Michael P. Johnson, who compiled a textbook containing a variety of case studies that can constitute "Community-Based Operations Research."

One of these case studies presented is the problem of park location in Bogotá, Colombia.

In the introduction I will explain the history of operations research and how it tends to be most utilized in the private sector. I will detail the development of (and necessity for) community-based operations research, and then explain that I will be examining a particular case study: the problem of developing new public parks in Bogotá, Colombia. I will perhaps provide a literature review here, a roadmap of what I will cover in the rest of the thesis, and anything else that may fit and come to mind later.

1.2 An Introduction to the Parks Location Problem

In urban areas and large cities, the presence of public parks, green spaces, and other recreational facilities has been associated with a significant improvement in quality of life, mental health, and general wellbeing [CITE SOURCE]. Knowing this, the city of Bogotá (Colombia), one of the largest cities in Latin America with a population of about eight million that is expected to reach ten million by 2025, has implemented a number of changes that include the recovery of public spaces and the improvement of public parks [CITE THE SOURCE].

In 2006, the mayor and city council of Bogotá threw their support behind a sports and recreation master plan for the city. This plan indicated that by 2019 the city must reach a minimum level of 2.71 m^2 of neighborhood park area per resident. It then became the *Instituto Distrital de Recreación y Deporte* (IDRD), or Recreation and Sports Institute of Bogotá's job to implement the master plan. As such, the IDRD was faced with a monumental challenge: they had to execute the construction of numerous new parks and revitalize dilapidated public spaces in a manner that balanced the differing geographic, social, and economic needs of the city.

Because of the many needs they had to consider and due to the nature of the problem, this problem was then modeled in such a way that it became a multiobjective facility location problem, which we will examine in further detail at a later point. For now we begin to consider the model.

Chapter 2

A Look at the Model

In this chapter I present a simplified version of the model. We will take a look at some simpler examples and gradually increase the complexity of our example until we arrive at the one presented in the paper.

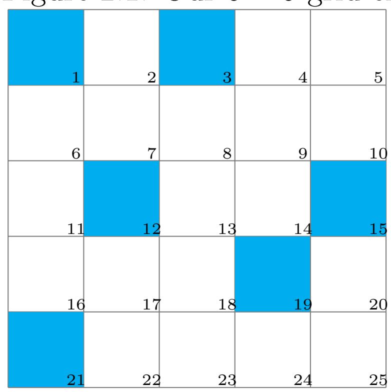
2.1 A Simplified Model

We first look at a model that has been adapted from a real model developed by researchers working in Bogotá. This simpler version of the model will help us understand the more complex project and can serve as a learning tool to both introduce the concept of multiobjective optimization and community based operations research.

2.1.1 Setting up the Problem

We begin with a 5×5 grid, which will represent our simplified fictional city. Each block represents a plot of land, which might either be empty or occupied. If it is occupied, this means that there are people living on the block. Otherwise, the lot is empty and it is a candidate parcel, to potentially be turned into a park. We number each block from one to 25, and we have colored every candidate parcel cyan. We want to build one park, and so we can choose from amongst all the cyan parcels. This is shown in Figure 2.1. We note that every lot is the same size. We also define the service area of the candidate parcels to consist of all the populated blocks adjacent and diagonal to the candidate parcel. For example, candidate parcel one's service area will

Figure 2.1: Our 5×5 grid city



consist of the set $\{2, 6, 7\}$. In other words, if we built a park on candidate parcel one, then the visitors would be from **only** those three surrounding lots. The first question we are interested in asking then becomes:

Which candidate parcel will maximize the geographical coverage, as measured by the service area of the parcel?

In this example, it is possible to answer our question after merely observing the grid. Candidate parcel 12 would result in the highest number of lots served with 8. We now move on to defining a certain number of sets and variables so that we can present a mathematical formulation of the problem.

We first let $\mathcal{J}$ be the set of all blocks that would benefit from the construction of a new park. Looking back to Figure 2.1, we see that $\mathcal{J} = \{2, 4, 6, 7, \ldots, 22, 23, 24, 25\}$. The candidate parcels have not been included in this set, and in this case we note that lot five is not included in $\mathcal{J}$ either because it is outside the service area of every candidate parcel. We also then define the set $\mathcal{I}$ to be the set of candidate parcels. In our case, $\mathcal{I} = \{1, 3, 12, 15, 19, 21\}$. We finally define the set $\mathcal{W}_j$ for $j \in \mathcal{J}$, which consists of all the candidate parcels that service block j. In our example, $\mathcal{W}_7 = \{1, 3, 12\}$.

We now define z_j as a binary decision variable, taking on the value 1 if block $j \in \mathcal{J}$ is covered by at least one park, and 0 otherwise. So if we decided to build a park on candidate parcel 21, then $z_{16} = z_{17} = z_{22} = 1$, and for every other possible value of j we would have $z_j = 0$. We also define the binary decision variable y_i, which will take on value 1 if candidate parcel i is selected to become a park, and 0 otherwise. If we turned candidate parcel 21 into a park, then $y_{21} = 1$. We can now formulate our initial model:

$$\begin{aligned}
\max\ f_1 &= \sum_{j \in \mathcal{J}} z_j \\
\text{subject to } z_j &\leq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
|\mathcal{W}_j|\, z_j &\geq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
p_{max} &\geq \sum_{i \in \mathcal{I}} y_i \\
z_j &\in \{0, 1\}, j \in \mathcal{J} \\
y_i &\in \{0, 1\}, i \in \mathcal{I}
\end{aligned}$$

Our objective function seeks to maximize the geographical coverage of the potential parks to be built. The first constraint guarantees that if block j is covered, then at least one parcel servicing it has been selected as a park. So if block 2 is covered in Figure 2.1, then either candidate parcel 1 or candidate parcel 3 should have been selected to become a park. Conversely, the second constraint guarantees that if block j is not covered, then none of the candidate parcels servicing it should be selected as parks. While I have referred to these first two constraints as being one individual constraint each, the observation can be made that they must be satisfied for all values of $j \in \mathcal{J}$, meaning that each constraint must be repeated $|\mathcal{J}|$ times.

The next constraint indicates how many parks we would like to build. The variable p_{max}, or the maximum allowable amount of parks we want built, is a constant that is decided upon beforehand. In our case, we decided that we wanted to build one park, so $p_{max} = 1$. We can think of this constraint as a simpler way to encode a budget into our model. Making the simplifying assumption that building a park on any lot will cost the same amount allows us to include this constraint in consideration of our budget. Should our budget change, we can just as easily change the value of p_{max}. The last two constraints define our z_j and y_i as binary decision variables.

Figure 2.2: Fully expanded model

$$
\begin{aligned}
\max\ f_1 = {} & z_2 + z_4 + z_6 + z_7 + z_8 + z_9 + z_{10} + z_{11} + z_{13} + z_{14} \\
& + z_{16} + z_{17} + z_{18} + z_{20} + z_{22} + z_{23} + z_{24} + z_{25}
\end{aligned}
$$

$$
\begin{aligned}
\text{subject to } z_2 &\le y_1 + y_3 & |W_2| z_2 &\ge y_1 + y_3 \\
z_4 &\le y_3 & |W_4| z_4 &\ge y_3 \\
z_6 &\le y_1 + y_{12} & |W_6| z_6 &\ge y_1 + y_{12} \\
z_7 &\le y_1 + y_3 + y_{12} & |W_7| z_7 &\ge y_1 + y_3 + y_{12} \\
z_8 &\le y_3 + y_{12} & |W_8| z_8 &\ge y_3 + y_{12} \\
z_9 &\le y_3 + y_{15} & |W_9| z_9 &\ge y_3 + y_{15} \\
z_{10} &\le y_{15} & |W_{10}| z_{10} &\ge y_{15} \\
z_{11} &\le y_{12} & |W_{11}| z_{11} &\ge y_{12} \\
z_{13} &\le y_{12} + y_{19} & |W_{13}| z_{13} &\ge y_{12} + y_{19} \\
z_{14} &\le y_{15} + y_{19} & |W_{14}| z_{14} &\ge y_{15} + y_{19} \\
z_{16} &\le y_{12} + y_{21} & |W_{16}| z_{16} &\ge y_{12} + y_{21} \\
z_{17} &\le y_{12} + y_{21} & |W_{17}| z_{17} &\ge y_{12} + y_{21} \\
z_{18} &\le y_{12} + y_{19} & |W_{18}| z_{18} &\ge y_{12} + y_{19} \\
z_{20} &\le y_{15} + y_{19} & |W_{20}| z_{20} &\ge y_{15} + y_{19} \\
z_{22} &\le y_{21} & |W_{22}| z_{22} &\ge y_{21} \\
z_{23} &\le y_{19} & |W_{23}| z_{23} &\ge y_{19} \\
z_{24} &\le y_{19} & |W_{24}| z_{24} &\ge y_{19} \\
z_{25} &\le y_{19} & |W_{25}| z_{25} &\ge y_{19}
\end{aligned}
$$

$$
\begin{aligned}
p_{max} &\ge y_1 + y_3 + y_{12} + y_{15} + y_{19} + y_{21} \\
z_2, z_4, z_6, z_7, z_8, z_9, z_{10}, z_{11}, z_{12}, z_{13}, z_{14}&, \\
z_{16}, z_{17}, z_{18}, z_{20}, z_{22}, z_{23}, z_{24}, z_{25} &\in \{0, 1\} \\
y_1, y_3, y_{12}, y_{15}, y_{19}, y_{21} &\in \{0, 1\}
\end{aligned}
$$

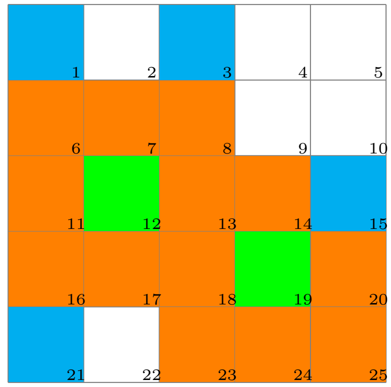


Figure 2.3: Picking two candidate parcels to maximize geographical coverage.

We can see an expansion of the model above to include all true constraints in Figure 2.2.

As we realized earlier, wanting to select only one candidate parcel will result in the selection of candidate parcel 12. What if we wanted to build two parks? Looking at Figure 2.3 we see that picking the two green parcels would result in a total coverage area of 13 blocks, highlighted in orange, which is the maximum in this case with two candidate parcels.

We were able to see this result without having to do any calculations because of the simplicity of our model. But this is not always so apparent. We now begin to alter our model and consider other objectives.

2.1.2 Changing the Objective

We were previously only concerned with maximizing the geographical coverage of our model. In real life, prioritizing the geographical coverage of the potential parks serves to spread out the location of the parks and avoids building excessive parks in densely populated areas. It also proactively locates parks in places that may have yet to be developed or may be susceptible to rapid population growth in the future. However, we do not want to fully neglect the population density of the various city blocks. How does our model change when we add the number of inhabitants in each block? The question we are interested in now becomes:

How do we maximize the geographical coverage of the parks as

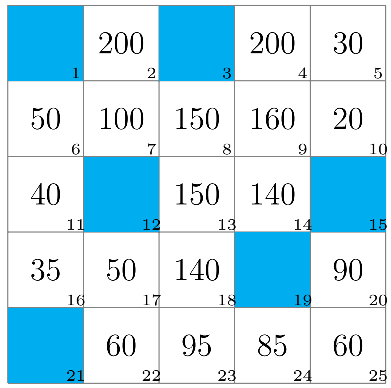


Figure 2.4: City blocks with population numbers.

well as the number of beneficiaries that would result from the construction of the parks?

We define the number of beneficiaries from a park construction as the sum of the population of all the blocks in the new park's service area.

We now turn to Figure 2.4 to see an updated version of our model city. The residential blocks have been updated with a number that indicates the number of people living in that city block. This allows us to calculate the number of beneficiaries that would result from the construction of a park on any candidate parcel. Building a park on candidate parcel 21, for example, would result in 145 beneficiaries. We define one additional variable.

Let p_i be the number of beneficiaries resulting from building a park on candidate parcel i. In the example of parcel 21, we would have $p_{21} = 145$. We can now update our model:

$$
\begin{aligned}
\max\ f_1 &= \sum_{j \in \mathcal{J}} z_j \\
\max\ f_3 &= \sum_{i \in \mathcal{I}} p_i y_i \\
\text{subject to } z_j &\leq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
|\mathcal{W}_j|\, z_j &\geq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
p_{max} &\geq \sum_{i \in \mathcal{I}} y_i \\
z_j &\in \{0, 1\}, j \in \mathcal{J} \\
y_i &\in \{0, 1\}, i \in \mathcal{I}
\end{aligned}
$$

Our model now has two objective functions. The first objective f_1 still seeks to maximize the geographical coverage of the parks. The second objective f_3 seeks to maximize the number of beneficiaries. But how do we maximize two things at the same time? We cannot guarantee that there will be a solution that maximizes the number of beneficiaries and the geographical coverage simultaneously. If we consider individual parcels, we notice that candidate parcel 3 serves the most amount of people, with $f_3 = 810$. However, candidate parcel 12 still has the most expansive geographical coverage, with $f_1 = 8$. How do we reconcile these two solutions?

2.2 A Solution Strategy

Because we are now trying to optimize multiple objectives, our question turns into a multiobjective optimization problem. We will present one possible way to solve these types of problems, borrowing the method used by the researchers from Bogotá. They implemented a lexicographic ordering of the objectives, which in practice meant that they ranked their objectives in accordance with their priorities. These priorities are reflected by the subscripts of the objective functions. Geographical coverage was deemed the most important criterion, which is why it is labeled f_1. Maximizing the number of beneficiaries was likewise deemed the third-most important criterion, so it is accordingly labeled f_3. For our model these are the only ones that we are concerned with, but it could work with any number of objective functions (for example, six).

The solution strategy is composed of two main parts. The first part is concerned with setting a benchmark for each objective function. The second part is about compromising to find a best possible solution. We will go over both parts in great detail.

2.2.1 Solving in Isolation

For the first part of our solution strategy, we will try to set a benchmark for each objective function by optimizing each of them in isolation. What does this mean? We return to our multiobjective optimization problem to explain the process.

$$
\begin{aligned}
\max\ f_1 &= \sum_{j \in \mathcal{J}} z_j \\
\max\ f_3 &= \sum_{i \in \mathcal{I}} p_i y_i \\
\text{subject to } z_j &\leq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
|\mathcal{W}_j|\, z_j &\geq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
p_{max} &\geq \sum_{i \in \mathcal{I}} y_i \\
z_j &\in \{0, 1\}, j \in \mathcal{J} \\
y_i &\in \{0, 1\}, i \in \mathcal{I}
\end{aligned}
$$

We see that our model has two objective functions, and we wish to solve for them in isolation. Our aim is to be able to treat each objective function as if it's the only objective function in our linear program. Since we have only two objective functions in this model, then one way to do this is by turning this model into two separate linear programs, as such:

$$\begin{aligned} \max\ f_1 = \sum_{j \in \mathcal{J}} z_j & \\ \text{subject to } z_j \leq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} & \\ |\mathcal{W}_j|\, z_j \geq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} & \\ p_{max} \geq \sum_{i \in \mathcal{I}} y_i & \\ z_j \in \{0, 1\}, j \in \mathcal{J} & \\ y_i \in \{0, 1\}, i \in \mathcal{I} & \end{aligned} \qquad \begin{aligned} \max\ f_3 = \sum_{i \in \mathcal{I}} p_i y_i & \\ \text{subject to } z_j \leq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} & \\ |\mathcal{W}_j|\, z_j \geq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} & \\ p_{max} \geq \sum_{i \in \mathcal{I}} y_i & \\ z_j \in \{0, 1\}, j \in \mathcal{J} & \\ y_i \in \{0, 1\}, i \in \mathcal{I} & \end{aligned}$$

Splitting our original multiobjective problem into multiple linear programs in this manner allows us to reduce our initial complicated problem into multiple, familiar, subproblems. We then solve each individual linear program, and we get the best possible result for each objective. Referring back to Figure 2.4, we again see that the first linear program will give us the result $f_1 = 8$, because that is the maximum geographical coverage possible. The second linear program will give us the result $f_3 = 810$, because as we established earlier, this is the maximum amount of beneficiaries served by any one candidate parcel. We will henceforth refer to these optimal values solved in isolation as f_1^* and f_3^*, with $f_1^* = 8$ and $f_3^* = 810$ as stated.

Performing this process, however, brings us no closer to finding a suitable solution for both objective functions simultaneously, because as we saw, each optimal solution for the linear programs results in different candidate parcels being picked (parcels twelve and three, respectively). What we have now, though, are figures that serve as an upper bound and a benchmark for our multiobjective problem. We know that, given no other considerations, when we want to build one park we can serve at most eight lots. Similarly, we know that when only considering the amount of people living in surrounding lots, we can serve at most 810 of them by building one park. In a problem that seeks to simultaneously maximize both of these figures, we can do no better than the individual optimal solutions we found.

In fact, this is a process that can be done for any number of objective functions. In our case, we are only working with two objectives, and so we split our program into two linear programs. With n objective functions, we would then accordingly split our multiobjective problem into n linear programs. We would then get an optimal individual solution for each objective, which would function as our best case scenario. In the final solution to our multiobjective problem, because we are trying to maximize both the geographical coverage and the number of beneficiaries served, then we want our solution to approach both optimal values, because as we noticed, we cannot satisfy both optimal values with the "budget" to build only one park.

We can now move onto the second stage of this process, where we present a methodical way to reconcile our solutions and arrive at an ideal middle ground.

2.2.2 Guess and Check

Before we dive into the second stage of the method, we must first define a term that will help us understand our process.

Definition 2.2.1 *Deterioration Rate For some value of the objective function f_k, the deterioration rate indicates how much worse our solution is than the optimal solution f_k^* as a percentage. In particular, we are interested in α_k, which will be the maximum allowable deterioration for objective k.*

Defining this idea of a deterioration rate now allows us to compare all of our solutions against the optimal solution derived from solving in isolation. For example, returning to our parks model, maximizing the geographic coverage alone resulted in a coverage of eight blocks. But maybe optimizing the population and the geographic coverage simultaneously might only result in a geographic coverage of six blocks. This six block solution would then be said to have a deterioration rate of 25%. If our solution for optimizing both the population and geographic coverage resulted in a solution with a geographic coverage of eight blocks, then our deterioration rate for the geographic coverage would be 0%. Finally, to see how every objective can have a deterioration rate, if our solution for the two objectives happened to be 5 blocks and 680 people respectively, then these solutions would indicate deterioration rates of 37.5% and 16%.

We can now define a value that is closely related to the deterioration rate.

Definition 2.2.2 *Compromise Threshold The compromise threshold indicates how "optimal" we want our solutions to be as a percentage of f_k^*. In particular, this is expressed as the value* $(1 - \alpha_k)$.

What this idea tells us is that for any solution that optimizes multiple objectives, we want it to give a value for f_k that is at least as good as some percent of f_k^*. If f_k was an objective that we really cared about, we might be more strict in demanding that any solution be at least 95% of the optimal value f_k^*. If we were less strict, we might be more lenient with our solution and say that it would only have to exceed 50% of the optimal solution f_k^*. This means that we can define different percentages for different objective functions according to how important we deem them to the problem at hand. Returning to our parks problem, if we wanted any compromise solution to satisfy at least 80% of the optimal value f_1^*, then our compromise threshold would be $1 - \alpha_k = 0.8$ and the maximum allowable deterioration would be $\alpha_k = 0.2$.

Why are these terms important? Well in fact, they are essential to stage two of our solution strategy, which I will explain first in general terms and then use our parks problem as a guiding example.

Remember that earlier, we had decided on ranking our objectives by order of importance. This is what the subscript in f_k denoted. Now, after we have solved for our objectives in isolation, we will need to solve some more linear programs to bring our many different optimal solutions together. We will start with the second-most important objective function, and modify the linear program that we solved in isolation for it. We will actually add a new constraint, which will look like the following: $f_1 \geq (1 - \alpha_1)f_1^*$. What does this mean? Remember that f_1 was our original most-important objective function. On the right hand side, we have our compromise threshold multiplied by the optimal value of the objective solved in isolation. So what this constraint does is ensure that the value of our objective does not drop below a certain target percentage of the optimal value. And in fact, this second stage of the process involves repeating this for every objective function. For the third-most important objective function and its linear program, we would add both $f_1 \geq (1 - \alpha_1)f_1^*$ and $f_2 \geq (1 - \alpha_2)f_2^*$ as objective functions. We would continue like this, moving on to the next linear program and turning the previous objective function into a constraint, until we get to our least important objective function, and every previous objective has been added as a constraint, along with the original constraints.

Returning to our parks problem to see an example of this strategy, we first note that we only have two objective functions, so our process will be much shorter. In fact, we have only one constraint, $f_1 \geq (1 - \alpha_1) f_1^*$, that we will add to the linear program with the third objective function. This constraint will actually look like this: $\sum_{j \in \mathcal{J}} z_j \geq (1 - \alpha_1)8$, because $f_1 = \sum_{j \in \mathcal{J}} z_j$ and $f_1^* = 8$. Remember that α_1 is our predetermined maximum acceptable deterioration rate. So our new linear program will look like this:

$$\begin{aligned}
\max\ f_3 &= \sum_{i \in \mathcal{I}} p_i y_i \\
\text{subject to } z_j &\leq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
|\mathcal{W}_j|\, z_j &\geq \sum_{i \in \mathcal{W}_j} y_i, j \in \mathcal{J} \\
p_{max} &\geq \sum_{i \in \mathcal{I}} y_i \\
\sum_{j \in \mathcal{J}} z_j &\geq (1 - \alpha_1)8 \\
z_j &\in \{0, 1\}, j \in \mathcal{J} \\
y_i &\in \{0, 1\}, i \in \mathcal{I}
\end{aligned}$$

To recap, in this linear program we try to optimize for the number of beneficiaries while ensuring that our geographical coverage does not drop below a certain percentage. And in fact, the solutions will change depending on what value for α_1 we decide to use. So how can we decide what value of α_1 will yield an ideal solution, and in general, how can we decide what values for any α_k to use? This is where automation will come in handy. Since it may not be particularly evident which value of α_1 will be the most optimal, we can employ a computer program to try many different values of α_1 for us. I devised a program to carry out precisely this task, and in Table 2.1 we can see the result of running this program. The table has four columns corresponding to the maximum allowable deterioration, the compromise threshold, the number of blocks served in the solution (the value for f_1), and the number

Allowable deterioration (α_1)	$(1 - \alpha_1)$	# blocks served (% of max)	# beneficiaries served (% of max)
0.10	0.90	8 (100)	715 (88.3)
0.15	0.85	7 (87.5)	760 (93.8)
0.20	0.80	7 (87.5)	760 (93.8)
0.25	0.75	7 (87.5)	760 (93.8)
0.30	0.70	7 (87.5)	760 (93.8)
0.35	0.65	7 (87.5)	760 (93.8)
0.40	0.60	5 (62.5)	810 (100)
0.45	0.55	5 (62.5)	810 (100)

Table 2.1: How changing the value of α_1 changes the optimal solution

of beneficiaries served (the value solved for f_3). We note that the last two columns have percentage values in parenthesis. These figures express the solution as a percentage of the optimal solutions f_k^* solved in isolation (8 for geographic coverage and 810 for beneficiaries). In our case of the small 5×5 town, we notice that there seem to be a set of three different, optimal solutions. When dealing with multiple objectives, however, our definition of optimal becomes a little bit muddled. Looking at the chart, we first note that the program tells us that if we relax the constraint on geographic coverage only a little bit by letting $\alpha_1 = 0.1$, the best solution is still to optimize the geographic coverage fully. This then results in 715 people being served by that selected candidate parcel. However, if we relax our constraint a little bit more and let $\alpha_1 = 0.15$, our solution changes. The number of blocks served drops to 7, which is not as good as our previous solution, but now the number of people served by the potential park jumps to 760. Relaxing the geographic constraint a significantly higher amount by letting $\alpha_1 = 0.4$ then results in another decrease in blocks serviced with five, but it has a corresponding jump of people serviced to 810, which is the maximum amount of people we could service by building one park. With this set of solutions, it seems to be the case that as we decrease our expectations of how many blocks we should service, we are able to service more people.

Since we don't actually have any more objective functions, then stage two of the process is over. We have no other function to repeat it for, and we cannot add f_3 as a constraint to a new linear program because we would have no new objective function to optimize. We run into another problem, however, as there seems to be no clear best solution out of the ones given. No one solution is better than the others in every single way, as there is always some tradeoff. This is where domain specific knowledge influences our decision, and it is ideal to consult the urban planners and other interested parties to see which solution is ideal to them. What is more valuable? Having one potential park serve 45 more people at the cost of one less block of coverage, or having a potential park serve 95 more people at the cost of three less blocks of coverage? Do we want to serve as large a geographic area as possible at significant cost to the number of beneficiaries served? It's quite hard to say, and often, there won't always be a clear-cut solution to multiobjective problems. What we can do is present a series of ideal solutions, and leave the experts and involved parties to pick the best one in accordance with their experience and criteria for selection.

2.2.3 A More Complicated Example

With our previous example, we were only looking at a small town with 25 total blocks, and a small number of candidate parcels. It turned out to be a very feasible problem to eye-ball, and it was possible to calculate the geographic coverage and population totals for every single candidate parcel and compare them that way. But we know that towns may be much much bigger. So what if we wanted to look at a bigger area of land? Perhaps a specific section of a town, or a city district? It would be near impossible to brute force our solutions for much bigger plots of land. This is where the program I designed comes into play, and it can solve much more difficult problems. Consider Figure 2.5, which we note is a much bigger example than our initial problem, and which has numerous candidate parcels to be considered. Now it is still possible to calculate the necessary information for each candidate parcel, given that there are only twenty-three of them, but it might be extremely difficult to decide on the best selection of parcels to optimize our objectives.

Instead, by automating our solution strategy, we are able to quickly find a solution for any value of α_k, the maximum allowable deterioration rate for an objective. In fact, the program performs both steps of the solution strategy, first by solving the corresponding linear program for each objective in isolation and then carrying out the incremental constraint addition for each corresponding objective. Returning to Figure 2.5, by solving for the objec-

Figure 2.5: A Daunting Example.

50	55		40	60	85	65	120	105	95	85		65	45	
	50	45	60	110	120	95	110		100	65	75	40	45	30
80	75	55	70	105	130	115		105	115	80	50	65	50	
165		65		65	80	95	110	105		70	55	45		25
225	200	70	75	100		100	105	120	105	55		20	25	15
250	120		85	105	100	110	70	80	70	65	50	15	5	10
200	100	180	155	125	105	80		95	75	40	45	35		15
250			145		235	100	80	135	95	70		50	45	20
300		265	205	235	505	500	245	175	80	85	65	30		25
305	255	275	215	225	400	405	410	500	305		45	50	40	30

tives in isolation we find out that when we are only able to build five parks we can at most serve 40 blocks, or alternately, at most serve 6690 beneficiaries. That is, $f_1^* = 40$ and $f_3^* = 6690$, as defined earlier.

It is also possible to calculate solutions within a given interval of α_k with any desired step amount. For example, in Table 2.2 we present the range of solutions (at five percent intervals) that allows up to 45% deterioration of α_1. We note that allowing for no deterioration whatsoever of the first objective results in only just under 70% of the optimal number of beneficiaries being served by our park selection, which may not necessarily be the best solution. We also see that allowing for 45% deterioration of the first objective results in the maximum number of beneficiaries served, and the increased deterioration of the first objective corresponds to the optimization of the other objective.

Looking at the rest of the solutions, it seems again that there is no clear-cut optimal solution out of all the provided solutions. We could eliminate some solutions, perhaps, like the ones corresponding to $\alpha_1 = 0.0$ and $\alpha_1 = 0.45$, because we note that the solution corresponding to $\alpha_1 = 0.05$ has little drop off in the number of blocks served while largely incrementing the number of beneficiaries served, with the reverse being true for $\alpha_1 = 0.40$ as it relates to $\alpha_1 = 0.45$. In loose terms, we like losing a little in one area while gaining a lot in another. Beyond that, however, it becomes a little more difficult and subjective to choose which solution is the best. How much more important is the geographic coverage than the number of beneficiaries resulting from the construction of our parks? To an eye unfamiliar with the

Allowable deterioration (α_1)	$(1-\alpha_1)$	# blocks served (% of max)	# beneficiaries served (% of max)
0.00	1.00	40 (100)	4605.0 (68.8)
0.05	0.95	38 (95.0)	5805.0 (86.8)
0.10	0.90	37 (92.5)	5855 (87.5)
0.15	0.85	35 (87.5)	6035 (90.2)
0.20	0.80	33 (82.5)	6345 (94.8)
0.25	0.75	30 (75.0)	6395 (95.6)
0.30	0.70	28 (70.0)	6575 (98.2)
0.35	0.65	28 (70.0)	6575 (98.2)
0.40	0.60	25 (62.5)	6640 (99.3)
0.45	0.55	22 (55.0)	6690 (100)

Table 2.2: A Range of Solutions for Deteriorations of the First Objective

situation, it might look like the solutions corresponding to $\alpha_1 = 0.10$ and $\alpha_1 = 0.15$ are possibly the best, because they optimize for both objectives quite nicely, but since we do not know if we should give both objectives equal importance, it is difficult to say for sure. Subjectivity plays a big role in deciding which solution to pick!

We now turn our attention to the real model used in Bogotá.

2.3 The Real Model

The model proposed by the researchers in Bogotá linked to a geographic information system (GIS) that would allow decision makers to interact with it to determine which candidate parcels become new parks. In this section we do not discuss any technical considerations but rather present the model in its entirety.

2.3.1 Objective Functions and Constraints

When considering candidate parcels for selection there are several aspects we are interested in:

- geographical coverage

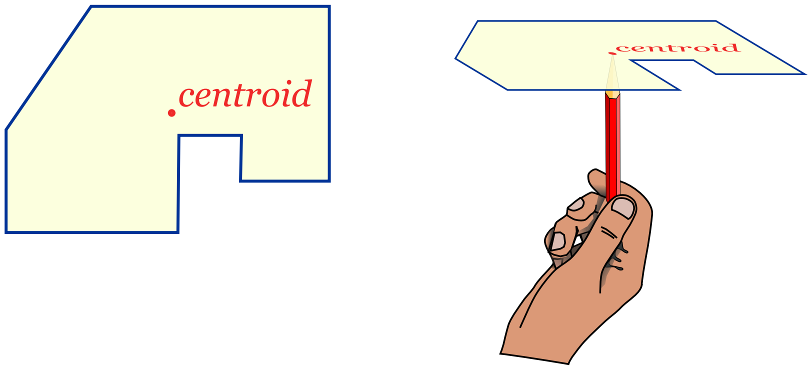


Figure 2.6: The centroid. [1]

- number of beneficiaries
- sidewalk and road accessibility
- nearby facilities
- cost

We have already interacted with the first two concerns, but in a real life scenario we must change how we define a candidate parcel's service area, thus affecting how we calculate these items as well. How do we define the service area then? One way is to draw a circular area of some radius from the parcel's centroid, which is essentially the center of mass of an object, as seen in Figure 2.6 (we can pretend that the shape is a cardboard outline of a candidate parcel). The radius from which the circular area is made can either be a fixed distance or it can be a function of the candidate parcel's area. That way, smaller parcels will have smaller service areas and larger parcels will similarly have bigger service areas. Using this new idea of the service area we can calculate the geographical coverage and the number of beneficiaries as before.

Continuing with the new indices we have the accessibility index which is based on the density of sidewalks and roads that make a certain parcel accessible. Given some candidate parcel's service area, the accessibility index is calculated as the total length of sidewalks and roads per square meter. This accessibility is then normalized by dividing it by the accessibility index of

[1] Image taken from MathIsFun.com.

the whole area of analysis, which means that a parcel with an accessibility index greater than one is relatively more accessible than the other parcels. Connectivity is an index that measures the connectedness between the candidate parcel and existing nearby facilities, which we can classify into one of two groups: negative and positive facilities[2]. So called negative facilities are ones that could potentially harm the perceived benefits of a park, like morgues and prisons, while positive facilities could be seen to increase the perceived benefits of a park, such as schools and recreational centers. The connectivity index is calculated by subtracting the number of negative facilities from the number of positive facilities within the parcel's service area. So an index greater than zero indicates that a park is expected to capture some extra benefits from the surrounding facilities. This index is purely a count, since there is no information that indicates the relative importance of the externalities provided by each facility, but it could be modified if such information were to become available by including a weight to reflect the importance of the effect of each facility.

There is also another aspect of considering nearby facilities which includes weighing them by their proximity. This perspective assumes that the impact of these facilities reduces as their distance from the parcel increases. The closer a facility is to a parcel, then the more it will affect it. As with the previous connectivity index, we subtract the negative facilities from the positive ones. Finally, the cost takes into consideration the potential cost of parcel acquisition as well as the construction costs of the park. This first component is estimated according to real estate appraisals while the second component is calculated based on construction materials, proposed park amenities, park design, operations, and further administrative obligations.

While it may not be immediately clear, these considerations are actually going to be the optimization objectives. The decision makers also ranked them in order of importance (a key component of the solution strategy), with the final order being the following:

1. Maximizing the geographic coverage

2. Maximizing the weighted proximity index

[2]While I will hereafter refer to facilities that may have perceived negative/positive effects for parks as "negative/positive facilities", this is by no means a condemnation of any urban facility (except for prisons, we should abolish those).

3. Maximizing the number of beneficiaries
4. Maximizing the accessibility index
5. Maximizing the connectivity index
6. Minimizing the cost

We will explore and further explain the mathematical formulations of the six objective functions in order to understand what is happening and then we will proceed to walk through the constraints so we can get a better picture of the entire model.

The goal of maximizing the geographic coverage is not too different than the one we presented in our simpler models. In fact, the only difference is that the definition of the service area has changed, as previously discussed. So for the real model we will still have the following as our objective function for the geographic coverage:

$$\max\ f_1 = \sum_{j \in \mathcal{J}} z_j \tag{2.1}$$

We move onto the second objective function which concerns the weighted proximity index. I will first present the mathematical formulation and then we walk through every component of it.

$$\max\ f_2 = \sum_{i \in \mathcal{I}} \left(\sum_{\{k \in \mathcal{E}_{\mathcal{P}} : d_{ik} \le r^e\}} \left(1 - \frac{d_{ik}}{r^e}\right) y_i - \sum_{\{k \in \mathcal{E}_{\mathcal{N}} : d_{ik} \le r^e\}} \left(1 - \frac{d_{ik}}{r^e}\right) y_i \right) \tag{2.2}$$

Clearly, there is a lot here for us to unpack. What is this objective function attempting to quantify? We are actually trying to measure the externality proximity index, or the impact (positive or negative) of facilities near the park. As described earlier, the further a facility is from a park the less of an effect it will have on said park. So a school across the street from a park will have a much greater impact on the park than a morgue a couple of blocks over. How does this function measure these things? We start by looking at the first term within the parenthesis of the function: $\sum_{\{k \in \mathcal{E}_{\mathcal{P}} : d_{ik} \le r^e\}} \left(1 - \frac{d_{ik}}{r^e}\right) y_i$. We start off by defining what certain sets mean. The set $\mathcal{I}$ is the set of all candidate parcels. The set $\mathcal{E}$ defines the set of all urban facilities in the area of analysis (in the town, or region we are considering for this model). Thus, the set $\mathcal{E}_\mathcal{P}$ refers to the set of facilities that would have positive effect on a park. The variable d_{ik} refers to the distance from the centroid of candidate parcel i to the facility k. The variable r^e defines the maximum distance of effect for any facility. Thus, the term $\left(1 - \frac{d_{ik}}{r^e}\right) y_i$ indicates as an index the effect from a facility on candidate parcel i. We note that the conditions under the summation symbol, $\{k \in \mathcal{E}_\mathcal{P} : d_{ik} \leq r^e\}$ indicate that we are only considering the positive facilities for this term, and only those whose distance to the centroid of the park is within their area of effect. Thus, the term $\sum_{\{k \in \mathcal{E}_\mathcal{P} : d_{ik} \leq r^e\}} \left(1 - \frac{d_{ik}}{r^e}\right) y_i$ indicates that we are measuring the effect of positive facilities on a candidate parcel. Similarly, the term being subtracted, $\sum_{\{k \in \mathcal{E}_\mathcal{N} : d_{ik} \leq r^e\}} \left(1 - \frac{d_{ik}}{r^e}\right) y_i$, indicates the effect of negative facilities on a candidate parcel. We note that in this term, we are considering facilities $k \in \mathcal{E}_\mathcal{N}$, where $\mathcal{E}_\mathcal{N}$ is the set of facilities with negative effect on parks. Thus, subtracting these two terms gives us an index of how positively or negatively affected a park is by the surrounding facilities. Finally, the final summation symbol indicates that we repeat this process for every single candidate parcel. And since we are trying to maximize this objective function, it means that we want to select parcels who will benefit the most from the surrounding facilities' presence.

We now move onto the third objective function, where we try to maximize the number of beneficiaries. Again, this is the same as in our model:

$$\max\ f_3 = \sum_{i \in \mathcal{I}} p_i y_i \tag{2.3}$$

For a refresher on what this means feel free to turn back to Section 2.1.2.

We now turn to the accessibility index, which is given by:

$$\max\ f_4 = \sum_{i \in \mathcal{I}} v_i y_i \tag{2.4}$$

In this objective function, the variable v_i represent the accessibility index of candidate parcel i. As discussed earlier, this is calculated by measuring the total length of sidewalks and roads within the service area of a candidate parcel, and then dividing this measure by the total length of sidewalks and roads over the *entire* area of analysis (the town, city neighborhood, etc). An index greater than one indicates a candidate parcel that is relatively more accessible than the rest. This is a figure that is calculated for all the candidate parcels, and is something we also try to maximize, because the more accessible a park is, the easier it is for people to go to it!

We now consider the connectivity index, which is expressed formulaically as:

$$\max\ f_5 = \sum_{i \in \mathcal{I}} e_i y_i \tag{2.5}$$

In this objective function, e_i denotes the connectivity index of parcel i. How is this calculated? By looking at the service area of a candidate parcel, we can determine all of the facilities within the service area that may have an impact (positive or negative) on the park. We then take all of the positive facilities within the service area of a parcel and subtract from this number the quantity of negative facilities. So for example, if there are five positive facilities and 3 negative facilities within a parcel's service area, then it would have a connectivity index of two. An index greater than zero might indicate that the park is expected to capture some extra benefits from the surrounding facilities. In fact, this index is related to the weighted proximity index but is slightly different, as it does not take into account the proximity of the facilities and ranks them all as equal, regardless of distance.

We can now finally consider the most important objective, where we consider the effect of park-building on our budget[3]. In this objective function, we wish to minimize the costs of our operations:

$$\min\ f_6 = \sum_{i \in \mathcal{I}} (c_i^l + c_i^b) y_i \tag{2.6}$$

We have two variables to consider here. The first one, c_i^l, indicates the parcel (or lot, hence the l) acquisition cost. The second one, c_i^b, indicates the cost of building (hence the b) a park on parcel i. These two variables are measured using real estate appraisals for the former, and material cost, park amenities and design, operations, and other administrative duties. Thankfully, this was ranked as the least important objective to fulfill. Saving money is clearly not the goal here, but if we could be given the choice between building two otherwise identical parks, we might as well choose the less expensive one. Thus

[3] This is a joke.

it is included as an objective to optimize, but clearly it is not the priority.

We can now begin to consider what sorts of constraints we might be bounded by. We will first present the constraints that we had already tackled in our simpler model.

$$z_j \leq \sum_{i \in \mathcal{W}_j} y_i \ , \quad j \in \mathcal{J} \tag{2.7}$$

$$|\mathcal{W}_j| \, z_j \geq \sum_{i \in \mathcal{W}_j} y_i \ , \quad j \in \mathcal{J} \tag{2.8}$$

These constraints guarantee that if a block j is covered, then some parcel covering the block has been selected, and conversely, if block j is not covered, then none of the parcels servicing it should be selected. As a quick refresher, the set $\mathcal{J}$ consists of all the blocks that could benefit from the construction of a park, z_j is a binary valuable indicating whether block j is being serviced or not, and the set W_j is the set of i that could service block j. For a more in-depth explanation, please refer back to Section 2.1.1.

This next constraint indicates that total new park area should meet a minimum requirement.

$$\sum_{i \in \mathcal{I}_{\mathcal{F}}} a_i y_i + \sum_{i \in \mathcal{I}_{\mathcal{V}}} x_i \geq \underline{a} \tag{2.9}$$

We note that with this constraint, two new sets are introduced: $\mathcal{I}_{\mathcal{F}}$ and $\mathcal{I}_{\mathcal{V}}$. The set $\mathcal{I}_{\mathcal{F}}$ contains the candidate parcels with an area less than or equal to $10000m^2$ (fixed-sized parcels, hence the $\mathcal{F}$ subscript). The other set contains larger parcels, which are referred to as variable-size parcels because the resulting parks built on these candidate parcels may not cover the entire area of the parcel (they might be smaller than the parcel). We note that the union of these two sets is $\mathcal{I}$, our original set. In other words, $\mathcal{I}_{\mathcal{F}} \cup \mathcal{I}_{\mathcal{V}} = \mathcal{I}$. The variable a_i denotes the area (in square meters) of parcel i. Since fixed-size parcels will cover the entire area of their parcel, then we can calculate the area for these resulting parks directly by using the parcel area. For variable sized parcels, however, we let x_i represent the area of candidate parcel i that is to be turned into a park. The variable $\underline{a}$ denotes the minimum requirement for park area as defined by Bogotá's recreational master plan (for a reminder of this plan, turn back to Section 1.2).

On the other hand, we also do not want to build too many parks, because then the IDRD (Recreation and Sports Institute) might be overwhelmed and incapable of administering the excess of parks. Thus, we also implement an upper bound on the amount of parkland we want built:

$$\sum_{i \in \mathcal{I}_{\mathcal{F}}} a_i y_i + \sum_{i \in \mathcal{I}_{\mathcal{V}}} x_i \leq \bar{a} \tag{2.10}$$

We note this constraint is almost identical to the previous one, with $\bar{a}$ denoting the upper bound for parkland to be built.

The next constraint puts a size limit on the parks built on variable sized parcels.

$$\underline{s}_i y_i \leq x_i \leq \bar{s}_i y_i \ , \quad i \in \mathcal{I}_{\mathcal{V}} \tag{2.11}$$

The variables $\underline{s}_i$ and $\bar{s}_i$ denote the minimum and maximum construction size parameters for the allotted park area in variable sized parcels.

We return to familiar constraints, with:

$$y_i \in \{0, 1\} \ , \quad i \in \mathcal{I} \tag{2.12}$$

$$z_j \in \{0, 1\} \ , \quad j \in \mathcal{J} \tag{2.13}$$

These constraints define our variables to be binary, as discussed in our simpler model.

Finally, our last constraint indicates that a variable sized park area cannot be negative (what is negative area?).

$$x_i \geq 0, i \in \mathcal{I}_{\mathcal{V}} \tag{2.14}$$

Put together, the grand model looks like this:

$$\max\ f_1 = \sum_{j\in\mathcal{J}} z_j$$

$$\max\ f_2 = \sum_{i\in\mathcal{I}} \left(\sum_{\{k\in\mathcal{E}_{\mathcal{P}}:d_{ik}\le r^e\}} \left(1-\frac{d_{ik}}{r^e}\right) y_i - \sum_{\{k\in\mathcal{E}_{\mathcal{N}}:d_{ik}\le r^e\}} \left(1-\frac{d_{ik}}{r^e}\right) y_i \right)$$

$$\max\ f_3 = \sum_{i\in\mathcal{I}} p_i y_i$$

$$\max\ f_4 = \sum_{i\in\mathcal{I}} v_i y_i$$

$$\max\ f_5 = \sum_{i\in\mathcal{I}} e_i y_i$$

$$\min\ f_6 = \sum_{i\in\mathcal{I}} (c_i^l + c_i^b) y_i$$

$$\text{subject to } z_j \le \sum_{i\in\mathcal{W}_j} y_i\ , \quad j\in\mathcal{J}$$

$$|\mathcal{W}_j|\, z_j \ge \sum_{i\in\mathcal{W}_j} y_i\ , \quad j\in\mathcal{J}$$

$$\sum_{i\in\mathcal{I}_{\mathcal{F}}} a_i y_i + \sum_{i\in\mathcal{I}_{\mathcal{V}}} x_i \ge \underline{a}$$

$$\sum_{i\in\mathcal{I}_{\mathcal{F}}} a_i y_i + \sum_{i\in\mathcal{I}_{\mathcal{V}}} x_i \le \bar{a}$$

$$\underline{s}_i y_i \le x_i \le \bar{s}_i y_i\ , \quad i\in\mathcal{I}_{\mathcal{V}}$$

$$y_i \in \{0,1\}\ , \quad i\in\mathcal{I}$$

$$z_j \in \{0,1\}\ , \quad j\in\mathcal{J}$$

$$x_i \ge 0\ , \qquad i\in\mathcal{I}_{\mathcal{V}}$$

Figure 2.7: The grand model

Chapter 3

Multiobjective Facility Location Problems

3.1 What?

Here we will look at the general class of Multiobjective facility location problems, which is the class of problems that contains the main model/problem I will be looking at for my thesis.

Chapter 4

Discrete Facility Location Problems

4.1 What? pt. II

The problem I will be examining also falls under this category. I am not sure how much this would vary from the previous chapter. But if it turns out there is a significant difference between both types of problems, it might be helpful to provide a dedicated chapter for both. Maybe the only difference is that the other class of problems has many objective functions.

Chapter 5

Applications

If I have time to apply what I have learned to another problem, maybe this will go here. But this may be a bit ambitious. Stay tuned.